The Dissertation Committee for Borghan Nezami Narajabad
certifies that this is the approved version of the following dissertation:

Essays on Dynamic Markets with Heterogeneous Agents

Committee:

Russell W. Cooper, Co-Supervisor

P. Dean Corbae, Co-Supervisor

Kenneth Hendricks

Randal B. Watson

Satyajit Chatterjee

Essays on Dynamic Markets with Heterogeneous Agents

by

Borghan Nezami Narajabad, B.S.,M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2007

To Asadollah and Parvin.

Acknowledgments

For the first chapter, I am heavily indebted to Russell Cooper and Dean Corbae for their guidance and support. I wish to thank Satyajit Chatterjee, Randal Watson, Tom Wiseman, Kenneth Hendricks and seminar participants at Federal Reserve Bank of New York, Brown University, University of Illinois at Urbana-Champaign, Federal Reserve Bank of Richmond, Southern Methodist University and Rice University for their helpful comments. The second chapter is a joint work with P. Dean Corbae. We wish to thank three anonymous referees for their comments. We also wish to thank Narayana Kocherlakota, Lars Ljungqvist, Andrei Shevchenko, Shouyong Shi, Max Stinchcombe, Ted Temzelides, Tom Wiseman, and especially Randy Wright for helpful comments on this paper. The third chapter is a joint work with Randall B. Watson. We are grateful for helpful comments to Kenneth Hendricks and Peter Howitt. Special thanks goes to Javad Yasari.

BORGHAN NEZAMI NARAJABAD

The University of Texas at Austin

May 2007

Essays on Dynamic Markets with Heterogeneous Agents

Publication No. ______________________

Borghan Nezami Narajabad, Ph.D.
The University of Texas at Austin, 2007

Co-Supervisors: Russell W. Cooper and P. Dean Corbae

During the past 25 years, household bankruptcy filings have quadrupled and the level of credit card debt has doubled. I try to explain both of these facts as a result of a more informative credit rating technology. I consider an environment where borrowers are heterogeneous with respect to their cost of default. Lenders have access to a rating technology which provides an exogenous signal about borrowers' default costs. As the signal becomes more informative, the credit market will provide a higher credit limit for borrowers with a high cost of default, hence allowing them to borrow more, which makes them more likely to default, while decreasing borrowing and default by those with a low cost of default. Using SMM, I estimate the model to match data on the averages of credit limit and debt as well as the increase in the spread of the credit limit distribution from the SCF 1992 and 1998. The model accounts for about one third of the increase in the number of bankruptcy filings.

In the second chapter we address Hotelling's venerable question about where shops endogenously locate in variety space in an environment that shares certain features of search models of money. Specifically, households are anonymous, have heterogeneous tastes, search is directed, and multilateral matching is possible. We solve for optimal incentive feasible allocations. We implement the solution as a trading post economy. In certain regions of the parameter space, the implementation shares features of a representative agent, cash-in-advance model while in other regions implementation of the optimum involves cross-sectional heterogeneity in consumption and production.

In the third chapter we use a computational methodology to incorporate endogenous horizontal differentiation into a dynamic stochastic model of quality investment under duopolistic competition. We analyze the effects on the industry's long-run innovation rate of changes in: (a) consumer transport costs, and (b) the costs of horizontal differentiation. More competition in the form of lower transport costs may have a U-shaped effect on innovation, rather than the inverse-U reported in earlier studies. The effect of less costly horizontal differentiation depends on the degree of consumer taste heterogeneity. Less costly differentiation raises innovation if taste heterogeneity is high, and lowers it otherwise.

Contents

Chapter 1

Information Technology and the Rise of Household Bankruptcy

Household bankruptcy filings have been increasing in the US for the past quarter of a century. In 1984, 0.33% of American households filed for bankruptcy. The number of filers rose to 0.93% of households in 1991 and continued to increase up to 1.41% in 2004.[1] This trend can also be spotted in the number of Canadian bankruptcy filers (Livshits et al. (2005)), suggesting that the increase should not be solely attributed to legal changes in the US.

During this period, households' access to unsecured credit (mainly through credit cards) flourished. While in 1989, 56% of households had access to credit cards and 29% of households carried a positive balance on their accounts. Fifteen years later credit card access rose to 72% and 40% of American households were carrying debt on their accounts (the latter are called *revolvers* in the literature).[2] Moreover, the average credit card debt of

[1] Just before the sweeping changes to America's bankruptcy code took effect in 2005, the number of bankruptcy filers jumped to 1.55% of American households. Unsurprisingly, the number of filers plummeted after the change went into effect. Recent data suggest the number of filers is picking up again.

[2] 20% of households (69 percentage of revolvers) were carrying more than \$500 debt in 1989. This fraction rose to 30% of households (75 percentage of revolvers) in 2004.

revolvers increased from $\$1,830$ in 1989 to $\$3,300$ in 2004.[3] But households were not just borrowing more subject to the same credit limits. During this period the average credit card limit available for an American household more than doubled; they rose from $\$7,100$ in 1989 to $\$15,200$ in 2004.[4]

The importance of credit card debt on a household's decision to file for bankruptcy has been well documented (see for example Domowitz and Sartain [22] as well as Sullivan, Warren and Westbrook [52].) Therefore, understanding the dynamics behind the expansion of credit card availability and its usage is critical for studying the rise of household bankruptcies.

Barron and Staten [8] document that expansion of the credit card industry would not be possible without rapid improvements in information technology and credit rating technologies. In 1997 credit bureaus issued some 600 million reports about credit seekers, (Padilla and Pagano [45]), and in the following decade credit scores produced by the Fair Isaac and Company, known as FICO scores, became the industry's standard tool for assessing borrowers' credit worthiness.

This paper tries to explain the rise in the number of bankruptcy filings as a result of improvement in the credit rating technology which allows the credit market to better screen borrowers' riskiness. This might sound counter intuitive at first. When creditors separate borrowers according to their riskiness, they will tighten the credit supply for the riskier borrowers, which will make them less likely to default. However, the safer borrowers will receive a higher borrowing limit which allows them to borrow more and, in turn, can result in more default. This is because even safer borrowers, ceteris paribus, are more likely to default when their debt level is higher. The net change of the debt level and the default rate is ambiguous.

Suppose the rating technology does not work well, so the credit market lacks information on borrowers riskiness and cannot differentiate them, which I call the "*pooling*

[3] All dollar amounts are in 1989 constant prices.

[4] A household's credit card limit is the sum of limits on all of the household's credit cards.

case". The equilibrium supply of credit will be so tight that the safer borrowers do not find it optimal to borrow much, and therefore are not paying much for the losses of the credit market from lending to the riskier borrowers.[5] Now, suppose the rating technology improves, so the credit market obtains information on the riskiness of borrowers and can differentiate them, which I call the "*separating case*". In this case, the market will cut back the supply of credit for the riskier borrowers only slightly. Hence the default rate by these borrowers does not fall very much. On the other hand, creditors will extend the supply of credit for the now distinguished safer borrowers extensively, encouraging them to borrow and hence default more.[6] I will call this the *informational* explanation for the rise of household bankruptcy.

The literature provides other explanations for the rise of household bankruptcies. The common explanation attributes it to the fall of "stigma" attached to bankruptcy. Gross and Souleles [25] report that ceteris paribus, a credit card holder in 1997 was almost 1 percentage point more likely to declare bankruptcy than a card holder with identical risk characteristics in 1995. Fay, Hurst and White [24] report that even after controlling for state and time fixed effects, households are more likely to file for bankruptcy if they live in districts with higher aggregate bankruptcy filings rates.

The stigma explanation, however, has counterfactual implications for credit availability and equilibrium debt levels. If borrowers become less reluctant to default on their debt, then shouldn't creditors restrain the supply of credit? and wouldn't this result in less borrowing rather than more? Athreya [4] and Livshits, MacGee and Tertilt [41] have noted that the decline in stigma alone would lead to a counterfactual decline in the ratio of unsecured debt to income. To account for the rise of consumer debt level they suggest a reduction in the transaction costs of lending.[7] Livhshits et al. [41] use a combination of decline in stigma and fall in transaction costs to explain the changes in filings and the ratio

[5]For example when the riskier borrowers are much more risky than the safer ones and there are enough of them in the pool of borrowers, borrowing and default will be mostly done by the riskier ones.

[6]Borrowers' responsiveness to the terms of credit contracts, and specifically credit limits, are well documented by Gross and Souleles [26].

[7]Athreya [4] also uses the same reduction to generate the rise in filings, which leads to a significantly higher debt to income ratio than that observed in the data.

of unsecured debt to income.[8]

But a fall in credit transaction costs cannot explain the increase in the spread of households credit card limits. While from 1992 to 1998, when bankruptcy filings rose significantly, the average American household's credit card limit increased from $\$7,200$ to $\$12,800$, the standard deviation of the cross sectional distribution of credit limits rose from $\$8,200$ in 1989 to $\$15,700$ in 2004.[9]. That is, the distribution of credit card limits did not just shift rightward, its spread also increased (the increase of credit limit for some households has been larger than the increase of limits for others). Gross and Souleles [25] report that creditors extended the larger lines to less risky accounts, suggesting that the spread of credit supply is mostly associated with the improvement in risk assessment.

Using a combination of the rise of stigma and fall of transaction costs to address different trends of the consumer credit industry ignores an important innovation of this industry; improvements of credit risk rating. This paper studies the implication of an improvement on the trends of the consumer credit industry.[10] Specifically, I examine how the credit limit and debt distribution as well as the number of bankruptcy filings differ in a market where creditors have information about the their borrowers' types from a credit market where they do not.

Section (2) provides some facts from the data on the trend of household bankruptcies, the distributions of credit card limits and debt as well as changes in these distributions across time. Section (3) describes a model of households' demand for credit, the responsiveness of their demand to credit contracts, in particular credit limits, and the response of propensity of default to an increase of credit supply. Then the model is used to show how a more informed credit market on average supplies more credit, but the distribution of credit

[8] Another possible explanation for more bankruptcy filings could be the rise of "uncertainty" in households' income and emergency expenses. This explanation implies a similar counterfactual decline in credit provision. Moreover, Livshits et al. [41] find its effect on the rise of filing numbers insignificant.

[9] Credit limits are reported in 1989 dollars

[10] Chatterjee et al. [16] provide a model with dynamic updating of creditors' beliefs about borrowers' creditworthiness that they associate with credit scores. Their paper does not, however, have anything to say about trends, which is the main point of my paper.

supply will also spread. Section (4) provides a simple quantitative example and shows the model does well in explaining the rise of credit supply, consumer debt level and the number of household bankruptcies. Section (5) concludes.

1.1 Data and Motivation

Households can file for bankruptcy under chapter 7 or 13. Under chapter 7 their unsecured debt such as credit card debt, installment loans, medical bills and damage claims are discharged, and filers lose all of their assets above an exemption level.[11] Under chapter 13, filers must propose a plan to repay a portion of their debts from future income without losing their assets. Since households have the right to choose between the chapters, they are only obliged to use future earnings to repay debt to the extent they would repay under chapter 7. Those who file under chapter 13 are allowed to file again under chapter 7, but the chapter 7 filers cannot file for another 6 years. The bankruptcy flag remains in a filer's credit history for 10 years (see Musto [44].)

Approximately 70% of those who seek bankruptcy protection file under chapter 7 and two third of those who file under chapter 13 ended up filing again under chapter 7.[12] This paper, however, does not distinguish between filing under the two chapters and studies a notion of bankruptcy similar to filing under chapter 7.

Figure(1.1) shows the number of bankruptcy filings by American households in the past two decades.[13] Except three short periods of 1992-94, 1997-2000 and 2003-04, bankruptcy filings have been increasing. From 1994 to 1997 bankruptcy filings increased by 63% during a period of robust economic expansion.[14] From the Survey of Consumer Finance (SCF), I find 11% of American households had at least one time filed for bankruptcy

[11] Exemption levels differ across states.

[12] See Li and Sarte [40] for an elaborated study of bankruptcy filers' choice of chapter.

[13] The percentage of filers for the 1984-95 period are reported from Fay et al. [24]. The number of filings for the 1995-2005 period are from www.uscourts.gov and the number of households for this period are from www.census.gov.

[14] Gross and Souleles [25] study bankruptcy and delinquency of credit card holders during this period.

in their lives by 2004, and from those who had filed 69.4% of them had filed in the past 10 years. That is, more than 7% of American households had a bankruptcy flag on their credit history.

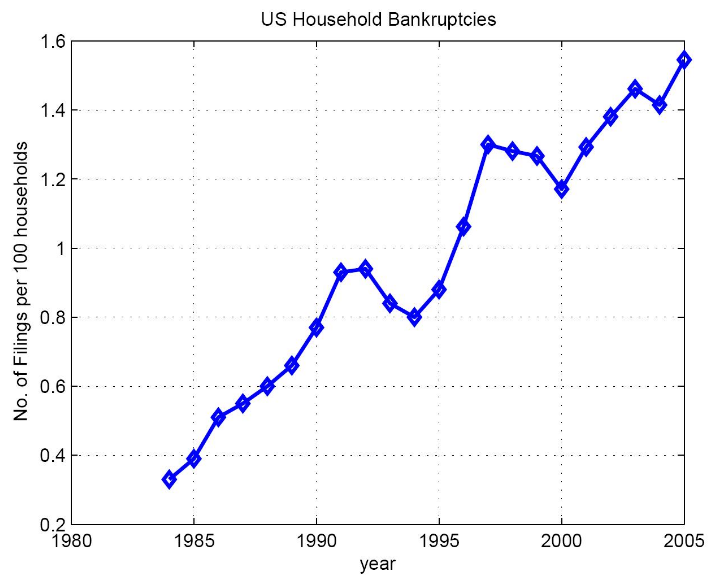


Figure 1.1: US Household Bankruptcies

Availability of credit cards and their usage has also been on the rise in the past two decades. Figure(1.2) and table (1.1) report the average credit limit for those who had access to credit cards and the fraction of population with credit card access from the SCF 1989-2004. The fraction of the population with positive credit card limit, which I call the *extensive margin* of credit supply, rose almost 17%. The average credit limit for card holders, which I call the *intensive margin* of credit supply, more than doubled.[15] Just from 1992 to 1998, the intensive margin of credit supply increased by a factor of 79%.

Households also borrowed more on their credit cards. In 1989, 29% of households were revolvers (carrying positive debt on their credit cards). By 2004 the fraction of revolvers rose to 40%. Revolvers' average credit card debt almost doubled in this period

[15]Limits are given in 1989 dollars.

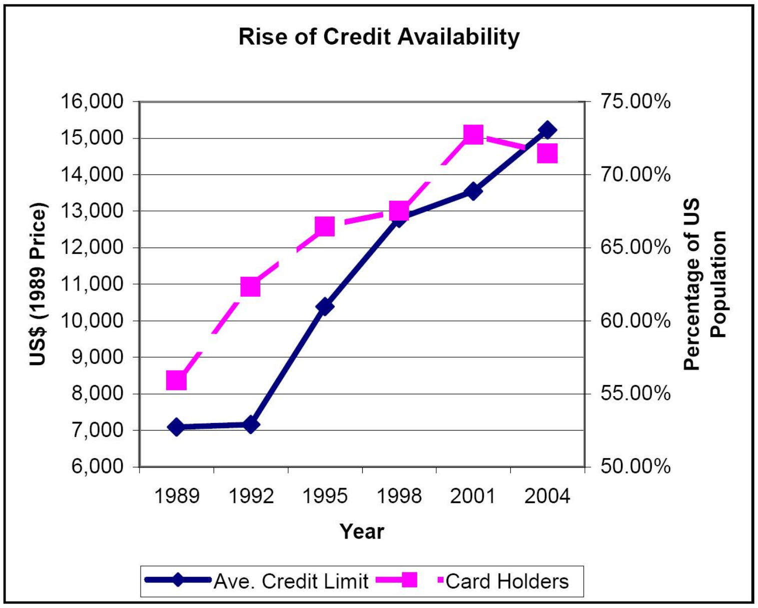


Figure 1.2: Rise Of Credit Availability

and went from $\$1,828$ in 1989 to $\$3,295$ in 2004. Just from 1992 to 1998, revolvers' debt increased by a factor of 59%. Table (1.1) reports the average debt level of revolvers and households with access to credit cards. The average debt level of revolvers remains almost two times as large as the average debt level of general card holders, revolvers and non-revolvers combined.

But the increase of average credit limits and debt levels does not thoroughly summarize the changes in the distributions of these two variables. The standard deviation of the cross section of credit limits and debt levels also doubled from 1989 to 2004. This observation is critical for the approach of the paper.

Figure (1.3) depicts the empirical distributions of credit limits (in 1989 dollars) in 1992 and 1998. As it can be easily noted, the distribution shifted rightward. But the shift was not caused by uniformed extension of credit supply to all card holders. The increase of credit limits for some households was larger than the increase of limits for others. To illuminate this point, figure(1.3) also depicts a counterfactual distribution, which is made

		1989	1992	1995	1998	2001	2004
Cred. Lim	Mean	7,092	7,157	10,390	12,802	13,548	15,223
	Debt > 0	7,125	6,579	9,832	11,505	11,964	13,643
Cred. Lim	Std.	11,296	8,223	13,151	17,861	22,055	20,911
	Debt > 0	9,624	7,204	11,233	15,696	21,645	18,066
Cred. Debt	Mean	954	1,025	1,346	1,696	1,453	1,851
(Card Holders)	Debt>0	1,828	1,947	2,404	3,098	2,707	3,295
Cred. Debt	Std.	2,120	2,303	3,076	3,979	4,172	4,246
(Card Holders)	Debt>0	2,648	2,878	3,788	4,958	5,390	5,228
Interest Rate	Mean	–	–	14.51	14.45	14.36	11.49
	Debt > 0	–	–	14.14	14.48	14.20	11.81
Interest Rate	Std.	–	–	4.29	4.63	5.24	6.42
	Debt > 0	–	–	4.46	5.04	5.62	6.63
Card Holders		55.91%	62.32%	66.45%	67.54%	72.72%	71.46%
Revolvers(Debt>0)		29.18%	32.83%	37.21%	36.97%	39.01%	40.14%

Table 1.1: Summary of US Households' Credit Cards

by uniformly increasing the credit limits in 1992 to match the average credit limit of 1998. Although the credit limit distribution of 1998 and the counterfactual distribution both have the same average, the 1998 distribution is more spread. The uneven extension of credit limits is also documented by Gross and Souleles ([25]). More interestingly they report that creditors extended the larger lines to less risky accounts and provided less extension to the riskier accounts.

Credit card contracts usually consist of a credit limit and an interest rate. Table (1.1) reports the average and standard deviation of credit card interest rates for 1995-2004.[16] From 1995 to 2001, when the Bank Prime Loan Rate (MPRIME) fluctuated between $8.00\% - 9.50\%$, the average of credit card interest rate remained around 14.5%, and its standard deviation rose almost one percentage point from 4.29% in 1995 to 5.24% in 2001.[17] The average of credit card interest rates in 2004 decreased to 11.49% while MPRIME dropped to $4.00\% - 5.00\%$. The variation of credit card interest rates across

[16]SCF did not collect the credit card interest rates prior to 1995.
[17]MPRIME is reported from the Board of Governors of the Federal Reserve System

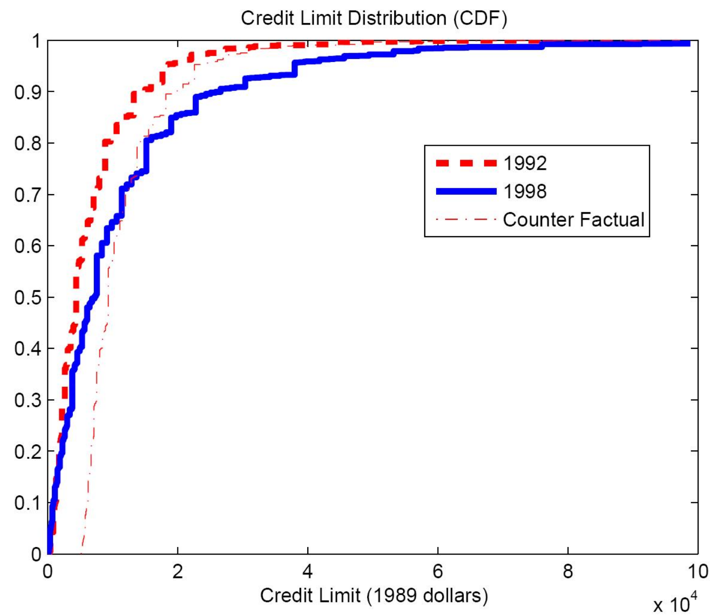


Figure 1.3: Distribution of Credit Limits

households also rose; specifically the standard deviation increased to 6.42%.

The simultaneous increase in the spreads of credit limits and interest rates indicates that creditors have started to offer more differentiated credit terms to their borrowers. Variation in credit limits, however, has increased far more than that of interest rates, especially prior to 2004. While the standard deviation of limits rose by a factor of 68% from 1995 to 2001, the standard deviation of interest rates increased by a factor of 22%.[18]. This facts motivates why I focus on changes in credit limits rather than variation in interest rates.

Gross and Souleles [26] study borrowers' response to credit supply and report an average "marginal propensity to consume (MPC) out of liquidity" (dDebt$/d$Limit) in the range of $10 - 14\%$. Their study finds that MPC is significant even for borrowers well below their limits. Average MPC of 14% implies a $\$790$ increase in the average credit debt for the $\$5,645$ increase of the average credit limit from 1992 to 1998. The actual average increase

[18]Stickiness of credit card interest rates have been studied by Ausubel [5] and Calem and Mester [13]

of debt level is $\$671$ for this period, suggesting that the rise of debt levels can be mostly attributed to the increase of credit supply.

According to Gross and Souleles [26] the long-term elasticity of debt to the interest rate is approximately -1.3. Although the SCF does not report interest rates for 1992, the implied change of the average debt level due to the change of interest rates from 1995 to 1998 is $\$56$, while the actual average debt level increased $\$2,412$. This fact again confirms the paper's approach of focusing on the quantity side of the supply of credit, namely credit limits, rather than the price of credit, namely interest rates.[19]

So far, I have reported the credit card limit and debt for an average household. But how about the credit card limit and debt level of those who file for bankruptcy? The households who report bankruptcy filing in the SCF, have usually finished their legal processes and their debt levels are discharged. Moreover, after filing for bankruptcy credit cards are cancelled so no information is available from the SCF on the filers' limit. Not being a panel dataset, the SCF does not allow me to observe the credit card limit and debt of households just before filing for bankruptcy.[20] Therefore, this paper uses the financial description of bankruptcy filers from Sullivan et al. [52].

Credit card debt(Ratio to income)	1991	1997
Mean	\$10,193(.531)	\$12,608(.767)
s.d.	\$13,751(.755)	\$15,380(1.154)
25th percentile	\$2,702(.122)	\$3,864(.167)
median	\$6,112(.310)	\$8,262(.469)
75th percentile	\$12,807(.645)	\$14,188(.874)

Table 1.2: Credit Card Debt Listed in Bankruptcy (in 1989 dollars) From Sullivan et al.

Table (1.2) reports the distribution of credit card debt listed in bankruptcy in 1991 and 1997 from Sullivan et al. [52]. Since the SCF data collection and report takes approx-

[19] Another challenge for studying the effect of interest rate on debt level lies in the fact that credit card prices contain other dimensions like cash back rates, flyer mileages and other point programs on which no data is available from the SCF.

[20] I tried to use the PSID, but credit limits are not reported in that dataset.

imately one year, the data on filers' credit card debt corresponds to 1992 and 1998 data from the SCF. The table also reports the ratio of credit card debt to income for bankruptcy filers. The average credit card debt of filers increased by a factor of 24%, and the median increased by a factor of 35%, suggesting that the distribution not only shifted rightward but also spread. The increase in the ratio of credit card debt to income is even higher. The average of the ratio of credit card debt to income for filers rose 44% while the median increased by 51%.

This data suggests those who were filing for bankruptcy in 1997 were defaulting on much higher levels of credit card debt. I will study how those who filed for bankruptcy could get access to more credit through higher limits and accumulate larger amounts of debt before defaulting on it.

The next section provides a framework to study the data facts I described in this section.

1.2 Model

Credit rating agencies usually use a borrower's credit history to assess her creditworthiness. Hence borrowers should potentially take into account the effect of their borrowing/payment decisions on their future terms of credit contracts. The natural method of modeling the credit market would be employing dynamic signaling models; these models, however, are very difficult to analyze.[21] This paper takes a simple approach to model the improvement of credit risk rating. Creditors receive a public signal about borrowers' types when credit contracts are made. I use more informative signals as a proxy for the improvement of the credit rating technology. This paper abstracts away from how credit scores are developed and just focuses on the information content of signals when a household starts borrowing

[21] For an example of a model with dynamic updating of creditors' beliefs about borrowers' creditworthiness see Chatterjee et al. [16]. Extension of credit over time in their model depends on the evolution of household credit scores (or Bayesian posteriors of household type). Effectively their contracts ration through price rather than quantity limits (which is the focus of my paper). As documented in the previous section, credit limits are an important part of the contract and have experienced the greater part of variation over time.

on its credit cards.

Musto [44] documents that creditors change their supply of credit as they lose information on the creditworthiness of borrowers due to the removal of the bankruptcy flag from their credit history ten years after the filing date. This paper tries to study what happens when creditors become more informed about the creditworthiness of borrowers due to a better credit rating technology.

It is important to notice the notion of safe and risky are relative. If a safe borrower (a borrower with a high cost of default) accumulates a large amount of debt, she may be more likely to default than a risky borrower (a borrower with a low cost of default) who has accumulated a small debt. The nature of costs associated with default are not the focus of this paper. These costs can have different pecuniary and non-pecuniary forms. An essential assumption of the paper is the heterogeneity of these costs across borrowers. That is different households with the same level of debt make different decisions on bankruptcy filings. Obviously as the level of the debt rises all households become more likely to default. This paper tries to study the implication of this heterogeneity for the supply of credit and bankruptcy filings.

As noted by Calem, Gorday and Mester [12] credit card balances show high persistence with yearly autocorrelation of 0.90. That is, households use credit cards for medium- or long-term financing rather than for short-term or unanticipated liquidity shocks. This fact is exploited by this paper in modeling households' motives for borrowing on their credit cards. Instead of using a Markovian income process with high persistence, I assume households start with an initial income, then at a random time they switch to their permanent income level, which is likely to be higher. In this framework, households are not using their credit lines to smooth their consumption whenever they receive a short-term income or liquidity shock. Instead, credit lines are used for long-term financing, an implication consistent with the data.[22] Moreover, this setup allows for a very simple link between the

[22]For an elaborate model of credit card usage with Markovian income process, see Chatterjee, Corbae, Nakajima and Rios-Rull [17]

supply of credit and households' debt level.

An average American household has about four credit cards, with usually different terms of contract (i.e. different credit limits and interest rates.) However, they tend to carry their debt on a single one which offers them the lowest interest rate. I assume each agent is only allowed to make credit contract with a single creditor she chooses. Without this assumption, the agent could be offered a continuum of contracts with incremental credit limits and increasing interest rates. In that case, the agent would start borrowing from the contract with the lowest interest rate and as her debt level increased she would use the contracts with higher interest rates. This approach is identical with offering the agent a menu of interest rates for different levels of debt.[23] In order to simplify the model and focus on the credit limit dimension of credit cards, I assume each agent can only choose one contract from the contract offers she receives from creditors.

The model is general in allowing the credit contracts to vary in both credit limit as well as interest rate. For the quantitative exercise provided afterward, however, I will assume all contracts have a fixed interest rate, and they only vary in the credit limit.

In the following subsections, first I describe the environment. Then the household's problem will be studied. The creditor's problem and existence of equilibrium will conclude the section.

1.2.1 Environment

Time is continuous and the horizon is infinite. The economy starts with a unit measure of agents denoted by $i \in (0, 1)$. Agents discount the future at rate β and their instantaneous utility from consumption is given by a strictly increasing and strictly concave function $u(\cdot)$. There is also a competitive market of risk neutral creditors with access to funds at rate $r \geq \beta$.

There is a *Rating Technology* which sends a public signal $\tilde{\theta}(i)$ about each agent's

[23] Chatterjee et al. [17] uses this approach.

risk type $\theta(i)$ at the beginning of the economy. Then each agent i realizes her type $\theta(i) \in [0, 1]$ which is private information. The joint distribution of types and signals, denoted by $\psi(\theta, \tilde{\theta})$ is public information.

Agents receive two streams of incomes. First, type θ agents draw their *initial* stream of income, y^I, from a type dependent distribution $F_\theta^I(\cdot)$. A type θ agent i continues with this stream of $y^I(i)$ units of income till a random switching time governed by a Poisson process with a type dependent parameter δ_θ. Once the switching time arrives, she will draw her *permanent* income, y^P, from a type dependent distribution $F_\theta^P(\cdot)$, and she will receive a certain stream of y^P units of income for the rest of her life. We assume the support of y^I and y^P are uniformly bounded away from zero for all types.[24] Moreover, assume $F_\theta^P(\cdot)$ does not have any mass point. Agent's incomes are publicly observable.

Agents can borrow from the credit market, but cannot save. Lending contracts can only be made at the beginning of the economy after receiving the public signals about agents types and before the agents realize their initial incomes. Agents are only allowed to contract with a single creditor from the pool of competitive creditors. Creditors are committed to their contracts with each agent till she realizes her permanent income at which point the contracts can be renegotiated in the competitive market. Lending contracts are constrained to have a fixed interest rate ρ and a credit limit L; that is their debt will accumulate at interest rate ρ and can increase up to L. After realization of y^P agents can make new credit contracts.

At any point in time, agents are allowed to exercise their option of bankruptcy. If an agent files for bankruptcy, all of her debt will be forgiven but she cannot borrow from the credit market anymore. Moreover, after filing for bankruptcy, agent i can only consume $\theta(i) \in [0, 1]$ fraction of her income from that time on.

To summarize, at the beginning of the economy a competitive credit market receives

[24]That is $\exists \epsilon > 0$ such that $F_\theta^I(\epsilon) = F_\theta^P(\epsilon) = 0 \; \forall \theta$.

a signal about each agent's type, and offers a credit contract which consists of a fixed interest rate and a credit limit. Then, the agents realize their type and initial income and start using their credit line until they realize their permanent income. At any point agents can default on their debt which will cause them to lose a fraction of their income for the rest of their life.

1.2.2 Agent's Problem

Given the offered credit contract, which we denote by the pair of credit limit and interest rate (L, ρ) and after realizing type, θ, and initial income, y^I, agents decide on how much to borrow/pay on their credit lines, and whether to file for bankruptcy or not. In particular, at the beginning of time an agent can choose $b(t)$, the amount of borrowing/payment on her credit line if she doesn't switch to her permanent income by time t, and whether to default on her debt at time t if she has not realized y^P by that time. Since agents can only default once, we can denote the time of filing for bankruptcy by T^*. That is, if the agent's income does not switch to y^P by time T^* she defaults at that time. Obviously agents can choose to not default on their debt before switching to permanent income, in which case T^* is set to infinity.

The sequential problem for a type θ agent who has realized income y^I and is offered contract (L, ρ) is given by:

$$V^I((L,\rho);(\theta,y^I)) = \max_{b(t),T^*} \left\{ \begin{array}{r} \int_0^{T^*} e^{-(\delta_\theta+\beta)t} \left[u(y^I + b(t)) + \delta_\theta V^P(D(t);\theta)\right] dt \\ \\ + \int_{T^*}^{\infty} e^{-(\delta_\theta+\beta)t} \left[u(\theta y^I) + \delta_\theta V^D(\theta)\right] dt \end{array} \right\} \tag{1.1}$$

where the debt level at time $t < T^* \leq \infty$, denoted by $D(t)$, must satisfy the credit limit constraint:

$$D(t) = \int_0^t e^{\rho(t-\tau)} b(\tau) d\tau \quad \leq \quad L \tag{1.2}$$

$V^P(D;\theta)$ is the expected value of realizing the permanent income for a type θ agent with debt level D, and $V^D(\theta)$ is the expected value of realizing the permanent income for a type θ agent who has defaulted on her debt and filed for bankruptcy before realizing her permanent income.

In particular, the expected value of realizing the permanent income after default is:

$$V^D(\theta) = \frac{1}{\beta} \int u(\theta y^P) dF_\theta^P(y^P). \tag{1.3}$$

When a type θ agent with D units of debt realizes her permanent income y^P, which is observable by the credit market, since $r \geq \beta$ she has no incentive to borrow from the credit market without the intention of defaulting on it. Hence the creditors will not allow her to use the remaining of her credit line after realizing her permanent income. Thus the agent has to decide on paying back her debt or to default on it. If the agent files for bankruptcy, the present value of her utility from consuming θ of the stream of her income is:

$$\frac{1}{\beta} u(\theta y^P).$$

If the agent decides to pay back her debt, since there is no uncertainty about her future income for the competitive credit market, the charged interest rate will be set at r. Then her problem is choosing the stream of payment amount p to maximize the present value of her consumption given by:

$$\max_p \int_0^\infty e^{-\beta t} u(y^P - p) dt,$$

subject to $\dot{D} = rD - p$. The Hamiltonian for this problem is given by:

$$\mathcal{H} = e^{-\beta t} u(y^P - p) + \lambda(rD - p)$$

which yields the solution $\dot{p} = (r - \beta)\frac{u'(y^P - p))}{u''(y^P - p))}$. In the case of $r = \beta$, the solution is given by $p = rD$, and hence the present value of the agent's utility is given by:

$$\frac{1}{\beta} u(y^P - rD).$$

In this case, if $y^P - rD \geq \theta y^P$ the agent will choose to consolidate her debt at interest rate r and pay it back, otherwise she will default on her debt and consume θy^P for the rest of her life. Therefore for $r = \beta$ we have:

$$V^P(D;\theta) = \frac{1}{\beta}\left[\int_0^{\frac{rD}{1-\theta}} u(\theta y^P) dF_\theta^P(y^P) + \int_{\frac{rD}{1-\theta}}^{\infty} u(y^P - rD) dF_\theta^P(y^P)\right]. \tag{1.4}$$

Agents have three decision to make: (i) whether to default or pay back their debt after realizing their permanent income, (ii) whether to default or not before realizing the permanent income, that is to set $T^* < \infty$ or $T^* = \infty$, and (iii) the sequence of borrowing/payment $b(t)$.

If agent decides to default before realizing her permanent income, she does not do so before using all of her available credit limit, or otherwise she can continue borrowing and default later.[25] Let's denote the time of reaching the credit limit by T. Then if the agent defaults before realizing her permanent income, that is $T^* < \infty$, then $T = T^*$. Later we will show under certain conditions that T is finite, that is even if $T^* = \infty$ and the agent does not default before switching to the permanent income, she will reach her credit limit in finite time if she does not realize her permanent income for that long.

[25] Notice that even if she realizes her permanent income the option of default is still available for her.

Let's denote the agent's borrowing at the time of reaching her credit limit by $b^* = b(T)$. If the agent does not default at the limit, she has to pay the interest charge of her debt to satisfy the credit limit, that is $b^* = -\rho L$. So if the agent does not default after reaching the credit limit the continuation value from (1.1) is equal to

$$\frac{1}{\delta_\theta + \beta}[u(y^I - \rho L) + \delta_\theta V^P(L;\theta)].$$

But if she defaults the continuation value will be

$$\frac{1}{\delta_\theta + \beta}[u(\theta y^I) + \delta_\theta V^D(\theta)].$$

So the agent defaults at the credit limit only if:

$$[u(\theta y^I) + \delta_\theta V^D(\theta)] \quad > \quad [u(y^I - \rho L) + \delta_\theta V^P(L;\theta)] \tag{1.5}$$

Now if (1.5) holds and the agent defaults at the limit, then $b^* = b(T) = b(T^*) \geq -\rho L$. In general:

Lemma 1 *Agent's borrowing/payment at the credit limit, $b^* = b(T) \geq -\rho L$, satisfies:*

$$u'(y^I + b^*) \quad \leq \quad \frac{[u(y^I + b^*) + \delta_\theta V^P(L;\theta)] - [u(\theta y^I) + \delta_\theta V^D(\theta)]}{\rho L + b^*} \tag{1.6}$$

with equality if $b^* > -\rho L$. Moreover, if (1.5) holds then (1.6) uniquely determines b^*.

Proof. See the Appendix. ■

In short, the left hand side of (1.6) is the marginal benefit of consumption for an agent just before she reaches her credit limit. The numerator of the right hand side is the difference between the stream of utility before and after default, and the denominator is the rate of debt increase (or equivalently the rate of approaching the limit as the time of default.) Overall the right hand side of (1.6) is the marginal cost of approaching the event of default due to increasing consumption.

Now knowing the agents decision at her credit limit and whether she defaults or not at the limit let's study her borrowing decision before reaching the limit. The Hamiltonian for the agent's problem (1.1) before reaching the limit is given by:

$$\mathcal{L} = e^{-(\delta_\theta+\beta)t}[u(y^I + b) + \delta_\theta V^P(D;\theta)] + \lambda[\rho D + b], \tag{1.7}$$

and the optimal solution must satisfy:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \rho D + b = \dot{D} \tag{1.8}$$

$$\frac{\partial \mathcal{L}}{\partial D} = e^{-(\delta_\theta+\beta)t}\delta_\theta V^P(D;\theta) + \rho\lambda = -\dot{\lambda} \tag{1.9}$$

$$\frac{\partial \mathcal{L}}{\partial b} = e^{-(\delta_\theta+\beta)t}u'(y^I + b) + \lambda = 0 \tag{1.10}$$

We have the solution for $b(T)$ from (1.6) which implies the $\lambda(T)$ from (1.10). Now solving for λ backward from (1.9) and then substituting it in (1.10) for $t < T$ we have:

$$\begin{aligned} u'(y^I + b(t)) \quad &= \quad e^{-(\delta_\theta+\beta-\rho)(T-t)}u'(y^I + b(T)) \\ &\quad - \int_t^T \delta_\theta e^{-(\delta_\theta+\beta-\rho)(\tau-t)} V_D^P(D(\tau);\theta)d\tau. \end{aligned} \tag{1.11}$$

The left hand side of (1.11) is the marginal utility from increasing debt level at time t. The right hand side gives us the two marginal costs associated with increasing debt level. The first expression is the marginal cost associated with getting closer to the limit, hence being credit constrained, and the second expression is the marginal cost of debt level if the agent realizes her permanent income before reaching the limit.

Figure (1.4) helps to see the marginal costs. If the agent increases her borrowing for a small period of time but does not alter it for the rest, then she will reach the credit limit sooner. This is depicted by the altered debt level reaching the credit limit sooner than the original debt level. The agent's consumption for the period of time just before when she used to reach the credit limit declines. On the other hand if the agent realizes her permanent income before reaching the limit, with the altered borrowing, she will carry more debt which is more costly to pay back.

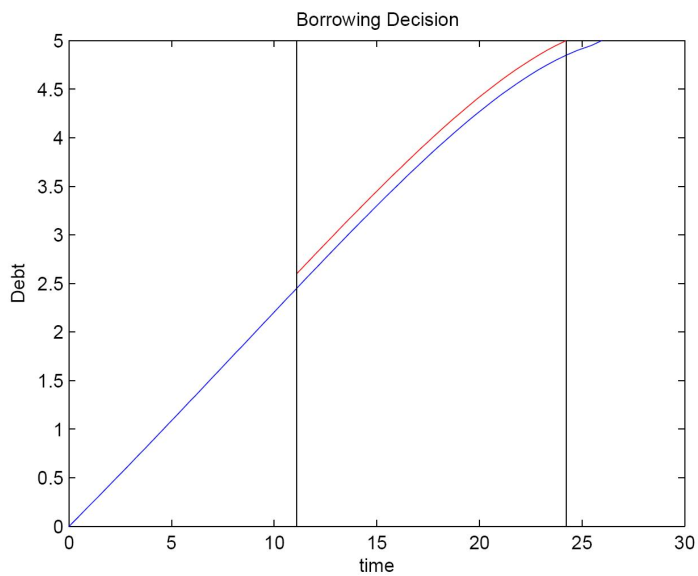


Figure 1.4: Change of Borrowing Decision

Notice that we can also consider the borrowing/payment amount, b, as a function of the outstanding debt, D, and the credit contract (L, ρ). That is $b(t) = b(D(t); (L, \rho))$. Then by taking the derivative of (1.10) with respect to time and then substituting for $\dot{\lambda}$ from (1.9) we have:

$$\frac{db(D; (L, \rho))}{dD} = \frac{\dot{b}}{\dot{D}} = \frac{(\delta_\theta + \beta - \rho)u'(y^I + b) + \delta_\theta V_D^P(D; \theta)}{u''(y^I + b)(\rho D + b)} \tag{1.12}$$

Using $b(L; (L, \rho)) = b^*$ from (1.6) as the boundary condition, we can find $b(D; (L, \rho))$ for $\forall D < L$ by solving the differential equation (1.12). Moreover, we can use this approach to state some properties of the borrowing/payment function.

Lemma 2 $b(D;(L,\rho))$ *is continuous in debt level,* D*, credit line,* L*, and interest rate* ρ*.*

Proof. By construction $b(L;(L,\rho)) = b^*$ is continuous from (1.6). The continuity of $b(D;(L,\rho))$ for $D < L$ follows from continuity of the solution for the differential equation (1.12). ■

Lemma 3 *For* $L_1 < L_2$*, if the agent's solution for (1.1) finds it optimal to increase the debt level up to the limit then:*

$$b(L_1;(L_1,\rho)) < b(L_1;(L_2,\rho)).$$

Proof. See the Appendix. ■

Theorem 4 *If the agent's solution for (1.1) finds it optimal to increase the debt level up to the limit then* $b(D;(L,\rho))$ *is strictly increasing in L.*

Proof. Suppose not, then $\exists D^*$ and $L_1 < L_2$, such that

$$b(D^*;(L_1,\rho)) \geq b(D^*;(L_2,\rho)).$$

Since $b(L_1;(L_1,\rho)) < b(L_1;(L_2,\rho))$ then by continuity $\exists\ D^{**}$ such that

$$b(D^{**};(L_1,\rho)) = b(D^{**};(L_2,\rho))$$

. But in that case (1.12) implies

$$b(D;(L_1,\rho)) = b(D;(L_2,\rho)) \quad \forall D \geq D^{**}$$

which contradicts $b(L_1;(L_1,\rho)) < b(L_1;(L_2,\rho))$. ■

The theorem state that when agents are offered higher credit limits they will accumulate more debt. This point can be seen in figure(1.5). With the credit limits depicted in this figure, agents do not default after reaching the credit limit and pay back the interest charge of their debt, waiting for the realization of their permanent income. Notice that with higher credit limit it takes longer for the agent to reach the limit. Moreover, as time passes and agents accumulate debt, their borrowing declines. Following (1.11) there are two factors contributing to the curbing of borrowing. First as the debt level goes up agents get closer to the credit limit which makes her borrowing constrained. Second as debt accumulates the marginal cost of debt after switching to the permanent income increases, hence making it more costly to borrow. Although the first factor is always in place, the second one may not.

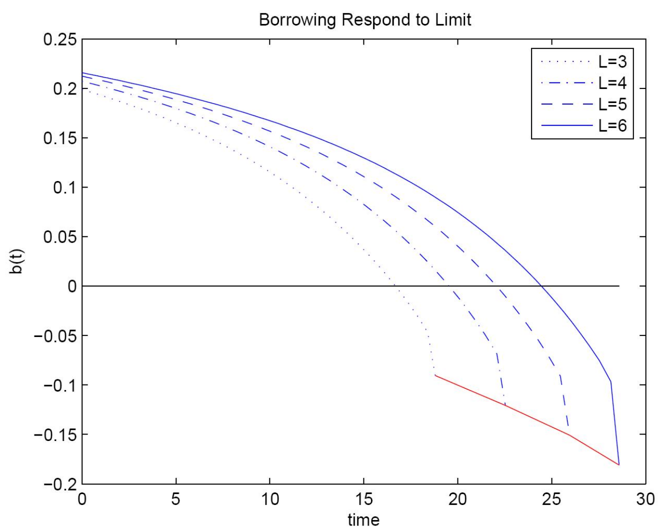


Figure 1.5: Borrowing with Different Credit Limits

Figure (1.6) shows the borrowing schedule for two close credit limits where with the lower one the agent does not default after reaching the limit but with the higher limit she finds the default option optimal once her debt level equals the credit limit.[26] This figure also depicts another important fact. Although eventually as the agent approaches her credit limit she curbs her borrowing due to the first factor described above, at the beginning as time passes and the debt level increase, the agent might actually increases her borrowing. This is because the second factor explained above is not in place. That is the marginal cost of debt after switching to the permanent income is actually shrinking.

Notice that as the debt level goes up, as long as the agent is not going to default on the debt after switching to her permanent income, the marginal cost of paying back also increases due to concavity of the utility function. But if the agent defaults after switching to the permanent income, then as the debt level rises the probability of paying back falls. Denote the lower bound for the support of y^P's distribution by $\underline{y}^P$. We can summarize this fact in the following lemma.

Lemma 5 *If $r = \beta$, then $V_D^P(D;\theta)$ is decreasing in D for $D \in [0, \frac{\underline{y}^P(1-\theta)}{r}]$ and increasing for $D > \frac{\underline{y}^P(1-\theta)}{r}$.*

Proof. See the Appendix. ■

Notice that so far we have simply assumed that the agent continues to accumulate debt untill reaching the credit limit. But what if the charged interest rate ρ is so high, or the agent does not expect very high permanent income, such that she does not find increasing the debt level up to the credit limit optimal? In particular if the solution for (1.12) is such that $\exists D < L$ where $b(D;(L,\rho)) = -\rho D$, then the agent stops borrowing after reaching to the debt level D. The following lemma provides a sufficient condition for optimality of borrowing up the credit limit.

Lemma 6 *For $r = \beta$, if $(\delta_\theta + \beta - \rho)u'(y^I) + \delta_\theta V_D^P(\frac{\underline{y}^P(1-\theta)}{r};\theta) > 0$ then the agent does*

[26] The figure shows that there are times when the agent borrows more with the lower credit limit. This is because the agent has accumulated much more debt by that time when she has a higher limit, and for $L_1 < L_2$ we still have

$$b(D;(L_1,\rho)) < b(D;(L_2,\rho))$$

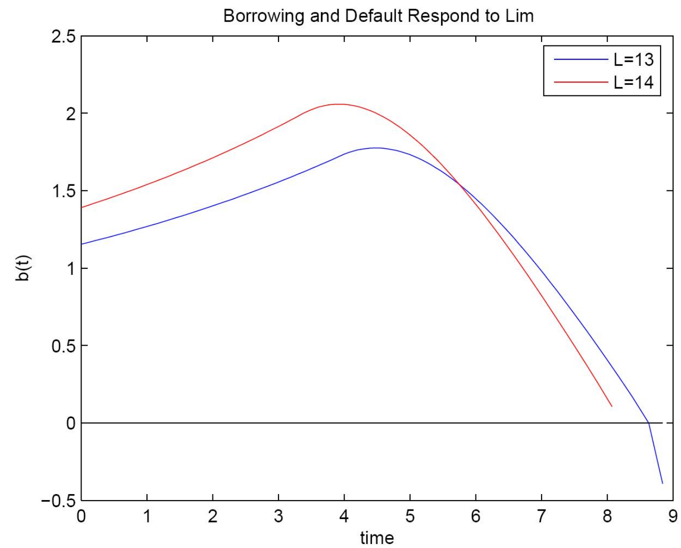


Figure 1.6: Borrowing and Default with Different Credit Limits

not stop increasing her debt before getting to the credit limit.

Proof. See the Appendix. ■

Notice that this condition is independent of the credit limit, L. However, it depends on the contract's interest rate, ρ. Obviously it also depends on the initial income as well as the agent's expectation of the time of realizing the future income, δ_θ, the distribution of permanent income, $F_\theta^P(\cdot)$, and the cost of default $(1-\theta)$. Although the previous lemma provides a sufficient condition for the agent to continue borrowing, it doesn't say that the agent actually reaches the limit or not, which is the subject of the next theorem.

Theorem 7 *For* $r=\beta$, *if* $(\delta_\theta+\beta-\rho)u'(y^I)+\delta_\theta V_D^P(\frac{y^P(1-\theta)}{r};\theta)>0$ *then there exists a time* T *such that for* $t\geq T$ *we have* $D(t)=L$. *That is, if agents do not realize their permanent income for a long enough time, then they reach the credit limit.*

Proof. Suppose the agent does not reach the credit limit, then for a large enough t such that

$D(t) \approx L$ and $b(t) \approx -\rho L$, (1.11) implies:

$$\begin{aligned} u'(y^I + b(t)) &= -\int_t^\infty \delta_\theta e^{-(\delta_\theta+\beta-\rho)(\tau-t)} V_D^P(D(\tau);\theta) d\tau \\ &\approx \frac{-\delta_\theta}{\delta_\theta+\beta-\rho} V_D^P(L;\theta) \end{aligned}$$

which contradicts $(\delta_\theta + \beta - \rho)u'(y^I - \rho L) + \delta_\theta V_D^P(L;\theta) > 0$. ■

So far, we have studied the effect of the credit limit on borrowing and debt levels, but not the default rate. From (1.5) the probability of default after switching to the permanent income with D units of debt is given by:

$$F_\theta^P(\frac{rD}{1-\theta}) \tag{1.13}$$

When the agent is offered a higher credit limit (as for example depicted in figure (1.5)) she will accumulated more debt. Since the time of switching to the permanent income is exogenous and independent of the debt level, after switching, the probability of having higher debt level and therefore the probability of default is higher with a higher credit limit.

Agents might also default when they reach their credit limit before switching to their permanent income. This is depicted in figure (1.6). In this comparison the higher credit limit not only causes the agent to accumulate more debt by the time of switching, but also makes the agent reach the credit limit sooner and then default after reaching the limit.

In summary, a larger credit limit induces agents to borrow more, and hence more likely to default. Ceteris paribus, an agent with low cost of default responds more to increase of credit limit and less to rise of interest rate, than an agent with high cost of default.

1.2.3 Creditors' Problem

In the previous subsection we studied decision rules of a type θ agent who has realized initial income y^I and is offered (L, ρ) credit contract. Creditors take agents' decision rules as given. Given the agents' decision rule as a function of the offered credit contract, creditors' expected profit from offering contract (L,ρ) to a type θ agent with initial income y^I is given by:

$$\Pi((\theta,y^I);(L,\rho)) \quad = \quad \int_0^{T^*} e^{-(\delta_\theta+r)t}[-b(t)+\delta_\theta D(t)(1-F_\theta^P(\frac{rD(t)}{1-\theta}))]dt \tag{1.14}$$

where $b(t)$ and T^* are the borrowing and default time decisions which solve (1.1) for a type θ agent with initial income y^I who is offered credit contract (L,ρ). Moreover $D(t)$ is the implied debt amount from (1.2). Notice that when the agent switches to her permanent income with debt level $D(t)$, the probability of paying back her debt is $1-F_\theta^P(\frac{rD(t)}{1-\theta})$ from (1.13).

When a creditor in the competitive credit market offers a credit contract, the only available information is the public signal $\tilde{\theta}$ from the Rating Technology. Agents have not realized their types nor their initial incomes yet. Denoting the conditional probability of drawing type θ given the signal $\tilde{\theta}$ by $\psi(\theta|\tilde{\theta})$, a creditor's expected profit from offering credit contract (L,ρ) to an agent with rating signal $\tilde{\theta}$ is:

$$\Pi(\tilde{\theta};(L,\rho)) \quad = \quad \int\int \Pi((\theta,y^I);(L,\rho))dF_\theta^I(y^I)d\psi(\theta|\tilde{\theta}). \tag{1.15}$$

Recall that when agents are offered the credit contract, they have not realized their type and income yet, however, they are also aware of the rating signal. Expected utility of an agent with the rating signal $\tilde{\theta}$ from contract (L,ρ) is given by:

$$V^I((L,\rho);\tilde{\theta}) \quad = \quad \int\int V^I((L,\rho);(\theta,y^I))dF_\theta^I(y^I)d\psi(\theta|\tilde{\theta}). \tag{1.16}$$

Since we assumed there is no asymmetry of information between agents and creditors when credit contracts are made, the only contract offered and accepted will be the one which delivers the highest expected utility subject to zero profit. That is the offered contract for an agent with rating signal $\tilde{\theta}$ must solve the following creditors' problem:

$$\max_{(L,\rho)} V^I((L,\rho);\tilde{\theta}) \tag{1.17}$$

$$s.t.$$

$$\Pi(\tilde{\theta};(L,\rho)) = 0.$$

Notice that when the signals are not very informative and hence agents with different types and incomes are lucked into the same contract, then some types generate positive profit which is used to compensate the losses made on other types. Moreover, lack of information causes another inefficiency: there are several pairs of (L,ρ) which can generate zero profit from a type θ agent with initial income y^I, but when agents with different characteristics are pooled and offered the same contract, they do not receive the efficient contracts.

Figure(1.7) shows how a creditor's profit changes as she increases the offered credit limit for a fix interest rate $\rho > r$. The figure depicts profit from two different types with the same initial income. For illustrative purpose I have assumed the two types have an identical distribution of permanent incomes and the same switching process, so their only difference is with respect to their default costs. As the limit increases the creditor's profit rises, since agents can borrow more but yet do not find it optimal to default on their debt, hence for $\rho > r$ the creditor's expected profit rises. For a large enough credit limit, the riskier agent accumulates enough debt to find it optimal to default, and creditors' profit from her falls. However, the safer agent still generates more profit. By increasing the credit limit eventually the safer agent also accumulates enough debt to find the default option optimal. This makes the expected profit from her to fall as well. Next section provides sufficient conditions under which solution for the creditor's problem (1.17) exists.

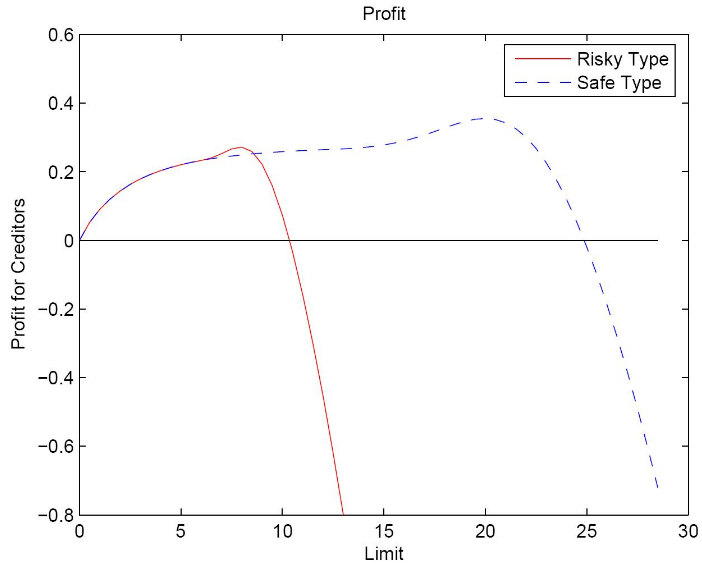


Figure 1.7: Creditors' Profit

1.2.4 Equilibrium Existence

In this part we show the existence of equilibrium, which is characterized by the solution for creditors' problem (1.17) of offering credit contract (L, ρ) to an agent with rating signal $\tilde{\theta}$.

Lemma 8 $\Pi((\theta, y^I); (L, \rho))$ *is continuous in* L *and* ρ*.*

Proof follows from the continuity of the decision rule b and lack of mass points in $F^P_\theta(\cdot)$[27].

Lemma 9 *If the support of* $\psi(\cdot|\tilde{\theta})$ *is bounded away from zero, the utility function* $u(\cdot)$ *is unbounded from above and the expected present value of all future income for an agent with signal* $\tilde{\theta}$ *is bounded then:*

$$\lim_{L\to\infty} \Pi(\tilde{\theta}; (L, \rho)) < 0 \ \forall \rho.$$

[27]When the supplied credit limit is such that the agent is indifferent between default or staying at the limit after reaching the credit limit, then creditor's profit depends on the fraction of agents who default after reaching the limit. Therefore $\Pi((\theta, y^I); (L, \rho))$ is a correspondence of L, however, a continuous one.

Proof. See the Appendix. ■

The proof first shows for any $\rho \geq r$ we have $\lim_{L\to\infty} V^I((L,\rho);\tilde{\theta}) \to \infty$. Then uses the fact that with $\Pi(\tilde{\theta};(L,\rho)) \geq 0$ the expected utility $V^I((L,\rho);\tilde{\theta})$ is bounded from above.

Theorem 10 *If the support of $\psi(\cdot|\tilde{\theta})$ is bounded away from zero, the utility function $u(\cdot)$ is unbounded from above and the expected present value of all future income for an agent with signal $\tilde{\theta}$ is bounded then creditors' problem (1.17) has solution.*

Proof. See the Appendix. ■

The idea of proof follows from defining $L^*(\tilde{\theta};\rho)$ as the largest L such that $\Pi(\tilde{\theta};(L,\rho)) = 0$. Then showing $V^I((L^*(\tilde{\theta};\rho),\rho);\tilde{\theta})$ attains its maximum.

1.3 Discussion and Quantitative Example

As stated in the previous section, when a contract is offered, neither the agent nor the creditor know the agent's true type. The only available information upon which a contract can be made contingent is the exogenous signal from the Rating technology. Therefore, contracts in my environment are only conditional on rating signals. That is, agents with identical signals will receive contracts with identical terms. However, after realization of types and initial incomes, agents with the same contract may choose different borrowing patterns. Moreover, the heterogeneous arrival and level of realized permanent incomes lead agents with identical contracts, type and initial income to accumulate different levels of debt and make different default decisions.

Although the model is general enough to allow both the credit limit and interest rate to be signal dependent, for the purpose of this discussion and the quantitative example, I make an assumption which is consistent with data. I assume credit contracts are all subject to a fixed interest rate $\rho > r$, and the only difference among types is in their cost of default.

In this case the only variation across contracts for different signals will be in their credit limits. Suppose different types are pooled together by the Rating technology, which is possible if the signals are not very informative about agents' types. In equilibrium agents will receive a credit limit which makes creditors' expected profit from the pooling contract equal to zero. The agents with a low cost of default will generate negative profit which will be compensated by the positive profit generated by the agents with high cost of default. Notice that since agents do not know their type at the time they choose from the offered contracts, and are not allowed to change their contract after realization of their type and initial income, at the time contracts are offered they have the same preference over the set of offered contracts. Therefore creditors cannot separate different types of agents within the pool of agents with identical Rating signals by offering different contracts.

Moreover, since the agents with high cost of default are more likely to pay back their debt, the marginal cost of borrowing is higher for them relative to the agents with a low cost of default. Therefore, with the same credit limit the high default cost agents accumulate lower levels of debt, and the responsiveness of their debt level to an increase of their credit limit is also smaller. This prevents creditors from raising the credit limit in the pooling case. Because if the credit limit is increased, then loans taken by low default cost agents which make negative expected profits will increase more than loans taken by high default cost agents which make positive expected profits.

Now suppose different types are separated by the Rating technology, which is possible if the signals are informative about agents' types. While in the pooling case there is cross subsidization across types, in the separating case the profit gain from high default cost type cannot be used to subsidize the loss made on low default cost type. In that case, the equilibrium supply of credit will increase significantly for the high default cost agents. Since the interest rate is assumed to be fixed, competitive creditors extend the credit limit for high default cost agents until a level where some of them increase their debt level high enough to find it optimal to default.

In the next subsection, I provide a quantitative example of the model to account for the increase in the average credit limit and credit card debt, as well as the rise in the number of bankruptcies observed from 1992 to 1998. In this quantitative example the mechanism which matches the data can be thought of as an increase in the intensive margin and not the extensive margin. Afterward, I will provide an explanation for the increase of extensive margin of credit supply.

1.3.1 Quantitative Example

Most of the households with positive credit limits do not borrow on their credit cards (see the last row of table (1.1)). Therefore I restrict my attention to revolvers, who have positive debt on their credit cards. The model predicts that households first increase their credit card debt before realizing their permanent income, then either default on their debt or pay it back. The SCF is not a panel data set so I could not use it to observe the dynamics of households' credit card debt accumulation or de-accumulation. The SCF, however, reports households' answer to the following question:

> Thinking only about Visa, Mastercard, Discover, Optima and store cards, do you almost always, sometimes, or hardly ever pay off the total balance owed on the account each month?

Roughly speaking, half of the revolvers answer "they hardly ever pay off the total balance." This group of households are those who are accumulating credit card debt and I call them *A-revolvers*. The last row of table (1.3) reports the fraction of A-revolvers from all households in the SCF 1992 and 1998. The first and third rows report average ratios of credit limit and credit card debt of A-revolvers to their annual income.[28] The standard deviations of the distributions of ratios of credit limits and credit card debt of A-revolvers

[28]This figures are reported after dropping less than half percentage of the subsample who report zero or negative income. I also tried the exercise with the average credit limit and credit card debt of A-revolvers divided by their average annual income. The results are very similar

to their incomes are also reported. The next row reports the fraction of A-revolvers who filed for bankruptcy, assuming all filers are also A-revolvers.

Since the fraction of A-revolvers in the population increased, looking at the ratio of defaulters to A-revolvers underestimates the rise of bankruptcy filings. However, the result of this quantitative exercise is not sensitive to using the ratio of defaulters to the average number of A-revolvers across two years.

A-revolvers	1992	1998
Av. (Cred Lim/Income)	25.97 %	42.48%
Std. (Cred Lim/Income)	41.38 %	77.50%
Av. (Cred Card Debt/Income)	10.66%	16.95%
Std. (Cred Card Debt/Income)	17.61%	35.46%
Default Rate	5.86%	7.39%
A-revolvers/Population	16.05%	17.31%

Table 1.3: Target Moments

For this exercise I assume the utility function to have constant relative risk aversion, that is $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$. The switching process from the initial income to permanent income follows a Poisson process with parameter δ for all types. The initial income is fixed for all types and the permanent income is drawn from a truncated exponential distribution:

$$F(y^P) = 1 - e^{-\eta(y^P - \underline{y^P})}$$

for all types. Finally I assume there are two types $0 < \theta_L < \theta_H < 1$ with distribution $\mu(\theta_L) = 1 - \mu(\theta_H)$.

In order to highlight the role of changes in the information technology, I will consider two cases. In the first one the signal from the Rating technology contains no information on agents' types so we have complete pooling of the two types. In the second case, I assume the signal is fully informative so we have complete separation and each type will receive a different contract.

I try to match the moments generated by the model in the pooling case to the data moments from 1992, and the moments generated by the model in the separating case to the data moments from 1998. Clearly the Rating technology provided some information on households' type in 1992, and did not provide full information in 1998. In particular, households were not offered a single credit card contract in 1992 as the model suggests in the pooling case. Instead, a distribution of credit card limits were supplied by the creditors in 1992. However, the spread of the distribution of credit limits increased from 1992 to 1998. In this quantitative exercise I try to account for the increase of the spread of the distribution of credit limits as a result of a more informative Rating technology[29].

In the pooling case, both types are offered a single credit limit L_P, while in the separating case two credit limits, $L_S^{\theta_L}$ and $L_S^{\theta_H}$, are offered. The 1992 data provides a distribution of credit limits. To study the increase of the spread of the credit limit distribution from 1992 to 1998, suppose any credit limit L from the distribution of credit limits in 1992 was replaced by two credit limits $\frac{L_S^{\theta_L}}{L_P}L$ and $\frac{L_S^{\theta_H}}{L_P}L$ with weights $\mu(\theta_L)$ and $\mu(\theta_H)$ for the distribution of credit limits for 1998. In this case, the following would hold for the coefficients of variation for these three distributions:

$$CV(L_{1998})^2 + 1 = (CV(L_{1992})^2 + 1)(CV(L_S)^2 + 1) \tag{1.18}$$

where L_{1992} and L_{1998} are the distribution of credit limits in 1992 and 1998, L_S is the distribution of credit limits in the separating case, and $CV(\cdot) = \frac{\sigma(\cdot)}{\mu(\cdot)}$ is the coefficient of variation.

Taking the time unit to be 3 months, I calibrate $\beta = r = \ln(.01)$ to be consistent with the 4% average annual growth rate. I set $\rho = \ln(.03)$ to be consistent with a 12%

[29] If the Rating technology sent signals about certain risk characteristics of borrowers in 1992, by 1998 the signals still contained the information about those characteristics. However, as the Rating technology became more informative, it could provide some additional information about the risk characteristics of borrowers. We can interpret the switch from the pooling case (uninformative signal) to the separating case (fully informative signal) as the provision of additional information by the Rating technology on borrowers.

interest charge on credit lines.[30] I set the coefficient of risk aversion to be equal to 1. I set y^I to be the average quarterly income of A-revolvers from the SCF which is $\$10,000$.[31]

β	ρ	σ	y^I
$\ln(0.01)$	$\ln(0.03)$	1	$\$10,000$

I estimate the three parameters related to permanent income plus the three parameters related to the default cost and its distribution to match six target moments. Four of the moments are the average ratios of credit limits to income and credit card debt to income for 1992 and 1998. The fifth target moment is the default rate of the A-revolvers in 1992. The last moment to match is the coefficient of variation of the credit limits in the separating case implied from 1992 and 1998 data by (1.18). The last target moment captures the increase in the spread of the credit limit distribution from 1992 to 1998. Notice that the exercise does not target the default rate of the A-revolvers in 1998.

An identity weight matrix is used to minimize the percentage deviation of the moments generated from the model from the targeted moments, which provides us with a consistent estimator. The estimated values are as follows and the moments generated from the model are reported in table (1.4) (the targeted values are inside the parentheses).

δ	$\underline{y}^P$	η	θ_H	θ_L	$\mu(\theta_H)$
0.1607	4.9635	0.0760	0.9893	0.9749	0.4251

The estimated parameters for permanent income implies on average households' permanent income is about 81% larger than their initial income, and on average it takes around one and a half years before switching to permanent income. Generally speaking this is consistent with the characteristics of the income process of households during financial

[30] Notice that the average credit card interest rate in this period is around $14.50 - 15.00\%$. However, since the model generates the equilibrium credit limits by equalizing creditors' profit to zero, we should consider creditors operational costs. I approximate this cost to be 3% from the difference between the Bank Prime Loan Rate and the Federal Fund Rate.

[31] Since I use $\log$ utility and the target moments are ratios relative to income, this variable is not important.

distress.

		Pooling	(1992)	Separating	(1998)
Target	Av. (Cred Lim/Income)	25.87%	(25.97 %)	42.92%	(42.48%)
Moments	Av. (CC Debt/Income)	10.65%	(10.66%)	16.74%	(16.95%)
	Default Rate	5.90%	(5.86%)		
	Coefficient of Variation			42.78%	(47.25%)
OverID	Default Rate			6.40%	(7.39%)
Moments	CC Debt/Income of Defaulter	10.51%	(53.10%)	40.40%	(76.70%)

Table 1.4: Generated Moments

I estimate six parameters to match six moments from the data. The default rate of 1998, however, is not amongst the target moments of this exercise and hence can be used to test for consistency of the model. According to data the default rate by A-revolvers rose from 5.86% in 1992 to 7.39% in 1998. The model generates an increase of default rate from 5.90% to 6.40%, which can account for about one third of the increase in bankruptcy filings in the data. Given the simple structure of the model and the fact that the exercise did not target the increase in the bankruptcy filings the result is quite appealing.

The ratio of credit card debt to income generated by the model and reported from data are provided in the last row of table(1.4) (Data moments are from Sullivan et al. [52]). The average credit card debt of a defaulter generated by the model is far less than the defaulters' credit card debt reported in the data. Moreover, although the model qualitatively matches the increase in the debt level of bankruptcy filers, it generates a far bigger increase than in the data.

Table(1.5) reports limits, debt, default rates and average debt level of a defaulter for both types in both cases. In the pooling case agents with high default costs (i.e. the safe type) borrow very little and do not default at all. Due to their small debt level they do not compensate the creditor a lot for the loss made on the riskier agents. Hence when they are separated borrowing and default by the riskier agents does not change significantly. However, in the separating case, the safer types are offered a much larger credit limit so they accumulate larger debts and this results in more frequent default.

		Pooling	Separating
Low Default Cost	Lim/Income	25.87%	21.57%
θ_H (Risky)	Debt/Income	15.58%	11.68%
	Def Rate	13.89%	7.62%
	Def Debt/Income	24.72%	20.88%
High Default Cost	Lim/Income	25.87%	58.71%
θ_L (Safe)	Debt/Income	7.01%	20.48%
	Def Rate	0.00%	5.50%
	Def Debt/Income	–	54.84%

Table 1.5: Generated Moments for Different Types

Although the model does relatively well in accounting for the rise of the default rate, it fails to quantitatively to match the increase of credit card debt levels of filers. This is because the safer types only default when their debt level is really high; therefore in the separating case the model generates a high level of credit card debt for the filers with the high cost of default.

Finally I use the estimated parameters to generate a counterfactual motivated by the stigma explanation of the rise of bankruptcy. Instead of changing the information structure from pooling to separation, I keep the pooling information structure but increase the fraction of high risk types to generate the increase of bankruptcy filings equal to the separating case. This goal is attained by changing the fraction of type θ_H from $\mu(\theta_H) = 42\%$ to $\mu(\theta_H) = 90\%$. However, this decreases the equilibrium credit limit to 21.98% of income and the average debt level increases only slightly to 11.53% of income, both contradicting the significant increasing trends observed in the data[32].

[32]To generate an increase in the default rate similar to the increase generated by the informational explanation, which is half of the increase observed in the data, $\mu(\theta_H) = 42\%$ should increase to $\mu(\theta_H) = 55.4\%$. In this case the equilibrium credit limit decreases to 24.28% of income, and the average debt level slightly increases to 10.92% of income, again contradicting the data trends.

1.3.2 Rise of Extensive Margin

The model can also explain the rise of the extensive margin of credit supply. As we noted earlier, the fraction of households with access to credit cards rose from 56% in 1989 to 72% in 2004. Assume there is a type who incurs no cost after default, that is $\theta = 1$. Supply of credit to this type only generates a loss for creditors and the loss is increasing in the credit limit. Now suppose a small fraction of another type $\theta^* < 1$, who incurs some cost from default and hence does not default on a low level of debt, is pooled together with type $\theta = 1$ by the Rating technology. In this pooling case, no credit will be supplied to the type θ^* agents who are pooled with type $\theta = 1$ agents. But when the Rating technology separates the two types, these type θ^* agents will receive a positive credit limit.

1.4 Conclusion

This chapter is a first attempt at providing an informational explanation for the rise of household bankruptcy. I simultaneously account for the increase of credit supply and the corresponding increase in average credit card debt. The extension of credit supply follows from the separation of revolvers with different degrees of riskiness. The rise of bankruptcy is explained by the increase in the availability of credit for the revolvers with high costs of default, which allows them to accumulate more credit card debt. While these revolvers are less likely to default on a low amount of debt, they default more frequently when they accumulate larger amounts of debt.

Using simulated method of moments I provide a simple quantitative example of matching the average credit limit and debt levels, as well as the increase in the spread of the credit limit distribution. The model accounts for about one third of the increase in the default rate, which is quite appealing given the simple structure of the model.

The model can be enriched by relaxing certain strong assumptions and can address other interesting questions. In particular it can be used to study why creditors find it optimal to vary the credit limit more than the interest rate. This may be informative about revolvers' income processes.

Understanding the rise of household bankruptcy has important policy implications. If the rise of bankruptcy filings is due to the decline of stigma, the policy response should be tightening the bankruptcy code to increase the cost of bankruptcy, similar to what the 2005 change of bankruptcy code tries to do. But if the rise of bankruptcy filings results from a more informed credit market, then tightening of bankruptcy code is not necessarily required.

Finally, Huggett [30] and Aiyagari [6] pointed out the importance of credit limits for households' precautionary saving motives and therefore the aggregate price of capital. This paper does not address household saving decisions, but tackles the question of how credit limits are allocated, which has an important role in household saving decisions and hence the aggregate capital stock. In that way, it can be used to complement the study of Chatterjee, et. al. [17] who study how bankruptcy affects the capital rate of return.

1.5 Appendix

Lemma 1 Agent's borrowing/payment at the credit limit, $b^* = b(T) \geq -\rho L$, satisfies:

$$u'(y^I + b^*) \quad \leq \quad \frac{[u(y^I + b^*) + \delta_\theta V^P(L;\theta)] - [u(\theta y^I) + \delta_\theta V^D(\theta)]}{\rho L + b^*}$$

with equality if $b^* > -\rho L$. Moreover, if (1.5) holds then (1.6) uniquely determines b^*.

Proof. If (1.5) does not hold, i.e. the agent does not find it optimal to default at the credit limit, then $b^* = -\rho L$ which satisfies (1.6). If (1.5) holds and the agent defaults at the limit, then (1.6) uniquely determines b^*. To see this point note that (1.6) can be rearranged as:

$$u'(y^I + b^*)(\rho L + b^*) - [u(y^I + b^*) + \delta_\theta V^P(L;\theta)] + [u(\theta y^I) + \delta_\theta V^D(\theta)] = 0. \quad (1.19)$$

If (1.6) holds, then left hand side of (1.19), which is decreasing in b^* due to concavity of $u(\cdot)$, is positive for $b^* = -\rho L$. Moreover, since $u(\cdot)$ is strictly concave as $b^* \to \infty$, left hand side approaches $-\infty$. Then by continuity (1.6) has a unique solution.

The optimality of solution for (1.19) follows from using calculus of variation for borrowing amount $\mathbf{b}$ when $D = L - \epsilon$ for a very small ϵ. Suppose the agent wants to maximize her utility for the next $\overline{\triangle t}$ periods, where $\overline{\triangle t}$ is small enough, after which she will for sure default. If the agent borrows a constant stream of $\mathbf{b}$ before reaching the credit limit, when she will default, then it approximately takes $\triangle t = \frac{L-D}{\rho D + \mathbf{b}}$ periods to reach the limit. So the agent will approximately receive utility

$$\triangle t[u(y^I + b^*) + \delta_\theta V^P(L;\theta)] + (\overline{\triangle t} - \triangle t)[u(\theta y^I) + \delta_\theta V^D(\theta)].$$

Taking first order condition with respect to $\mathbf{b}$ and then letting $\epsilon \to 0$, yields (1.19). ■

Lemma 3 For $L_1 < L_2$, if the agent's solution for (1.1) finds it optimal to increase the debt level up to the limit then:

$$b(L_1; (L_1, \rho)) < b(L_1; (L_2, \rho)).$$

Proof. If $b(L_1; (L_1, \rho)) = -\rho L_1$, that is the agent does not default at the limit, then since the agent increases her debt level up to the limit we should have $b(L_1; (L_2, \rho)) > -\rho L_1$.

If $b(L_1; (L_1, \rho)) = -\rho L_1$, that is the agent defaults at credit limit L_1, then from (1.5) it follows that she also defaults at credit limit L_2, therefore $b(L_1; (L_1, \rho))$ and $b(L_2; (L_2, \rho))$ are both governed by (1.19). From (1.19) it follows that:

$$\frac{db^*}{dL} = \frac{-\rho u'(y^I + b) + \delta_\theta V_D^P(D;\theta)}{u''(y^I + b)(\rho D + b)}. \tag{1.20}$$

Now comparing (1.20) and (1.12) it follows that

$$\frac{db^*}{dL} > \left.\frac{db(D; (L, \rho))}{dD}\right|_{D=L}$$

hence by continuity of b^* and $b(D; (L, \rho)$ it follows

$$b(L_1; (L_1, \rho)) < b(L_1; (L_2, \rho)).$$

■

Lemma 5 If $r = \beta$, then $V_D^P(D;\theta)$ is decreasing in D for $D \in [0, \frac{\underline{y}^P(1-\theta)}{r}]$ and increasing for $D > \frac{\underline{y}^P(1-\theta)}{r}$.

Proof. For $D \in [0, \frac{\underline{y}^P(1-\theta)}{r}]$, concavity of $V^P(D;\theta)$ follows from concavity of $u(\cdot)$ since the agent does not default after realizing her permanent income.

For $\frac{y^P(1-\theta)}{r} \leq D_1 < D_2$ we have:

$$\begin{aligned} V_D^P(D_2;\theta) &= -\int_{\frac{rD_2}{1-\theta}}^{\infty} u'(y^P - rD_2)dF_\theta^P(y^P) \\ &= -\int_{\frac{rD_2}{1-\theta}-r(D_2-D_1)}^{\infty} u'(y^P - rD_1)dF_\theta^P(y^P) \\ &> -\int_{\frac{rD_1}{1-\theta}}^{\infty} u'(y^P - rD_1)dF_\theta^P(y^P) \\ &= V_D^P(D_1;(y^I,\theta)) \end{aligned}$$

where the inequality follows from $D_2 > D_1$ and $u'(\cdot) > 0$. $\blacksquare$

Lemma 6 For $r = \beta$, if $(\delta_\theta + \beta - \rho)u'(y^I) + \delta_\theta V_D^P(\frac{y^P(1-\theta)}{r};\theta) > 0$ then the agent does not stop increasing her debt before getting to the credit limit.

Proof. The previous lemma guarantees that $\frac{\delta_\theta+\beta-\rho}{\delta_\theta}u'(y^I) + V_D^P(D;\theta) > 0$ for $\forall D > 0$.

Suppose the agent stops increasing her debt level above $\overline{D}$, that is $\rho\overline{D} + b(\overline{D}) = 0$, while $\overline{D} < L$. In this case the marginal benefit from increasing the borrowing amount is $(\delta_\theta + \beta)u'(y^I + b(\overline{D}))$ and the marginal cost is $\rho u'(y^I + b(\overline{D})) - \delta V_D^P(\overline{D};\theta)$. Concavity of $u(\cdot)$ guarantees $(\delta_\theta + \beta - \rho)u'(y^I + b(\overline{D})) + \delta V_D^P(\overline{D};\theta) > 0$ hence it is optimal to increase $b(\overline{D})$ and therefore the debt level. $\blacksquare$

Lemma 9 If the support of $\psi(\cdot|\tilde{\theta})$ is bounded away from zero, the utility function $u(\cdot)$ is unbounded from above and the expected present value of all future income for an agent with signal $\tilde{\theta}$ is bounded then:

$$\lim_{L\to\infty} \Pi(\tilde{\theta};(L,\rho)) < 0 \ \forall\rho.$$

Proof. First we show for any $\rho \geq r$ we have $\lim_{L\to\infty} V^I((L,\rho);\tilde{\theta}) \to \infty$. Next we show if $\Pi(\tilde{\theta};(L,\rho)) \geq 0$ then $V^I((L,\rho);\tilde{\theta})$ is bounded from above.

For any credit limit L, consider the plan of borrowing $b = \frac{\rho L}{e^{\rho}-1}$ during $t \in [0, 1]$, and then defaulting. Also always defaulting after realizing the permanent income. This plan delivers the expected present value of utility equal to

$$\int\int \frac{1}{\delta_\theta + \beta}\left[(1 - e^{-(\delta_\theta+\beta)})u(y^I + \frac{\rho L}{e^{\rho} - 1}) + e^{-(\delta_\theta+\beta)}u(\theta y^I) + \delta_\theta V^D(\theta)\right] dF^I_\theta(y^I)d\psi(\theta|\tilde{\theta})$$

This is a lower bound for $V^I((L, \rho); \tilde{\theta})$, and If the support of $\psi(\cdot|\tilde{\theta})$ is bounded away from zero, this lower bound goes to infinity as $L \to \infty$, due to unboundedness of $u(\cdot)$. Hence $V^I((L, \rho); \tilde{\theta}))$ is unbounded as $L \to \infty$.

Let's $\overline{y_{\tilde{\theta}}}$ denotes the stream of income which has the same present value as the expected present value of all future income for an agent with rating signal $\tilde{\theta}$ at interest rate r. That is

$$\overline{y_{\tilde{\theta}}} = \int\int\int \frac{1}{r + \delta_\theta}(ry^I + \delta_\theta y^P)dF^P_\theta(y^P)dF^I_\theta(y^I)d\psi(\theta|\tilde{\theta}).$$

Due to the concavity of $u(\cdot)$, if $\Pi(\tilde{\theta}; (L, \rho)) \geq 0$ then $V^I((L, \rho); \tilde{\theta}) \leq \frac{1}{\beta}u(\overline{y_{\tilde{\theta}}})$. Since $V^I((L, \rho); \tilde{\theta})$ is bounded from above when $\Pi(\tilde{\theta}; (L, \rho)) \geq 0$ and is unbounded when $L \to \infty$, we conclude $\lim_{L\to\infty} \Pi(\tilde{\theta}; (L, \rho)) < 0$ ■

Theorem 10 If the support of $\psi(\cdot|\tilde{\theta})$ is bounded away from zero, the utility function $u(\cdot)$ is unbounded from above and the expected present value of all future income for an agent with signal $\tilde{\theta}$ is bounded then creditors' problem (1.17) has solution.

Proof. By definition, for a given interest rate ρ, agent's expected utility $\Pi(\tilde{\theta}; (L, \rho))$ is increasing, as agents can opt out and do not use credit limit. Let's $L^*(\tilde{\theta}; \rho)$ denotes the largest L such that $\Pi(\tilde{\theta}; (L, \rho)) = 0$. Since $\Pi(\tilde{\theta}; (0, \rho)) = 0$ and $\Pi(\tilde{\theta}; (L, \rho))$ is continuous in L, the previous lemma guarantees the existence and uniqueness of $L^*(\tilde{\theta}; \rho)$. Creditors' problem (1.17) can be rewritten as:

$$\max_{\rho} V^I((L^*(\tilde{\theta};\rho),\rho);\tilde{\theta}) \tag{1.21}$$

Notice that $V^I((L^*(\tilde{\theta};\rho),\rho);\tilde{\theta})$ is continuous in ρ.

Following the proof of the previous lemma we know $V^I((L^*(\tilde{\theta};\rho),\rho);\tilde{\theta}) \leq \frac{1}{\beta}u(\overline{y_{\tilde{\theta}}})$ for $\forall\rho$. Since the support of y^I and y^P are uniformly bounded away from zero, the support of $\psi(\cdot|\tilde{\theta})$ is bounded away from zero and $\overline{y_{\tilde{\theta}}}$ is finite, $\exists\overline{\rho}$ such that for $\rho > \overline{\rho}$ no agent increases her debt level above zero if she is offered a contract with interest rate ρ, that is

$$V^I((L^*(\tilde{\theta};\rho),\rho);\tilde{\theta}) = V^I(0,r);\tilde{\theta}) \;\; \forall\rho > \overline{\rho}.$$

The bounded continuous function $V^I((L^*(\tilde{\theta};\rho),\rho);\tilde{\theta})$ attains its maximum on a compact set. ■

Chapter 2

A Hotelling Model with Money

2.1 Introduction

One question we address in this chapter is whether monetary exchange promotes product variety. The framework we use to answer this question is a dynamic version of a Hotelling [28] model. Hotelling's paper was the first to consider where firms would choose to locate in the product space given that consumers have heterogeneous preferences over a continuum of differentiated goods. Here we consider how the exchange mechanism interacts with preferences and technology to determine product variety.

Besides the relation to Hotelling's work, our paper is related to several other strands of literature. Starting with the insights of Kiyotaki and Wright [35] about the essentiality of fiat money in bilateral matching environments with double coincidence problems, there is a literature which studies agents' choices over which goods they specialize in production (see also Shi [50] and Camera, Reed, and Waller [14]). These papers show that the introduction of money can lead to more specialization relative to barter and can increase welfare in a bilateral random matching environments. There are many differences between our work and theirs. One difference, at the level of the environment, is that we allow multilateral matching and directed search.[1] Another difference, in terms of outcomes, is that despite increased specialization with money, the entire exogenous set of varieties of goods are produced in their framework while in our paper, since it is costly to set up a shop to exchange goods, there is a finite set of active shops and where they decide to locate in the variety space is endogenously determined. Another closely related paper is by Burdett and Shevchenko [11]. They endogenously determine the product space in a bilateral random matching framework with indivisible money. The indivisibility of money in their model means that product variety is not neutral with respect to the level of the money supply; a lump sum doubling of the level of money affects matching probabilities and product variety in their model rather than a simple doubling of prices (as in our framework).

We analyse this question by studying a planner's problem as in Kocherlakota [36] and then implement the allocation as a trading post economy. Shapley and Shubik [49] was one of the first papers to study exchange in a trading post economy. Recently, Howitt [30] extended this framework to one with possibilities of both barter and monetary trading posts and established conditions under which monetary exchange was preferred to barter.[2] Our paper differs from other trading post models by starting from the planner's problem and then considering implementation. This allows us to focus on the "optimal" product variety allocation and avoids all the problems associated with arbitrary beliefs and trading rules as well as coordination issues which have come up in the previous trading post models. Beliefs re-enter at the implementation stage and we show how differences in beliefs can lead to multiple equilibria with a pareto dominated outcome that can be eliminated through monetary policy.

We proceed as follows. Section II outlines the environment. Section III states the planner's problem and characterizes its solution. Section IV provides three illuminating ex-

[1] In this sense, our paper is closer to that of Laing, Li, and Wang [38]. However, there are big differences in our approach relative to theirs. At the level of the environment, for instance, all agents have identical "love of variety" preferences in their framework, which is quite different from the Hotelling or Kiyotaki-Wright environments with heterogeneous preferences.

[2] Another related paper which deals with endogenizing the exchange pattern in a trading post environment is Starr and Stinchcombe [51].

amples: how money may increase product variety, how the costs of operating trading posts may affect the choice of medium of exchange, and that there is a complementarity associated with opening money trading posts (specifically that there can be positive feedback from opening money shops while there is negative feedback from opening barter shops). Section V shows how to implement the allocation as a monetary trading post economy and studies implications of variations in the money growth rate for existence of equilibrium. We show that if the cost of operating shops is sufficiently low, households have sufficiently specialized preferences and are patient, and the money growth rate is not too high, the implementation shares features of a representative agent, cash-in-advance economy. In particular, while there is heterogeneity in the varieties of goods produced and consumed across households, households are symmetric with respect to the quantities of goods produced and consumed as well as money holdings. If these conditions do not hold, the planner's solution (and implementation) may entail cross-sectional variation in production, consumption, and money holdings. We end with an example where there are multiple pareto ranked equilibria due to beliefs about which trading posts will be opened. We show that positive money growth with linear prices may actually eliminate the inferior equilibrium while the implementation of the planner's allocation continues to exist.

2.2 Environment

Time is discrete $t = 0, 1, 2, ...$ The economy is composed of a unit measure of infinitely-lived households indexed by $h \in [0, 1]$ and a countably infinite number of infinitely-lived shopkeepers indexed by $k \in \mathbb{N}$. A household is composed of a shopper-worker pair.

Each good is indexed by its variety $\nu \in [0, 1]$. We let good 0 denote fiat money. Fiat money is storable, divisible, and generates no direct utility to households or shopkeepers. The quantity of money held by a household at the beginning of period t is denoted $m(t) \in \mathbb{R}_+$ and the quantity held by shopkeeper k is denoted $M^k(t) \in \mathbb{R}_+$. The money stock per capita at period t is given by $\overline{m}(t) = (1 + \gamma)^t \overline{m}(0)$ where $\gamma \geq 0$. Money is distributed to households via lump sum transfers $\eta(t)$.[3]

All goods other than money are nonstorable and consumable. Let $\mathcal{V} = (0, 1]$ be the set of such consumable goods, which are also divisible. Households are heterogeneous with respect to their tastes in much the same way as Kiyotaki and Wright [35]. Specifically, a household of type $\theta \in \Theta = \{(\tau_c, \tau_p) \in \mathcal{V} \times \mathcal{V}\}$ derives utility from consuming goods of variety ν_c in the interval $|\nu_c - \tau_c|_{mod1} \leq \frac{x}{2}$ and can produce goods of variety ν_p in the interval $|\nu_p - \tau_p|_{mod1} \leq \frac{x}{2}$ where $|\cdot|_{mod1}$ is a function which is $\min\{|\cdot|, 1 - |\cdot|\}$. As in Kiyotaki and Wright [35], we make the assumption that a household cannot consume its own output. The unit measure of households are uniformly distributed over the type space Θ. We assume the per period utility function for a type θ household associated with consuming $c \geq 0$ units of good ν_c and producing $\ell \in \{0, 1\}$ units of good ν_p can be written[4]

$$U_\theta((c, \nu_c), (\ell, \nu_p)) = W_\theta(c, \nu_c) + W_\theta(\ell, \nu_p)$$

where

$$W_\theta(c, \nu_c) = \begin{cases} c & \text{if } |\nu_c - \tau_c|_{mod1} \leq \frac{x}{2} \\ 0 & \text{otherwise} \end{cases}$$

and

$$W_\theta(\ell, \nu_p) = \begin{cases} -\varepsilon & \text{if } \ell = 1 \text{ and } |\nu_p - \tau_p|_{mod1} \leq \frac{x}{2} \\ -\infty & \text{if } \ell = 1 \text{ and } |\nu_p - \tau_p|_{mod1} > \frac{x}{2} \\ 0 & \text{if } \ell = 0 \end{cases}.$$

If the shopper stays home, $W_\theta(c, \nu_c) = 0$ and if the worker stays home, $W_\theta(\ell, \nu_p) = 0$. These preferences are simply a "donut" (two-dimensional) version of the (one-dimensional) consumption taste arc length x of the unit circle in Kiyotaki and Wright [35]. Households discount the future at rate $\beta < 1$.

[3] We do not consider $\gamma < 0$ since lump sum taxes could exceed household money balances on and off-the-equilibrium path.

[4] Caplin and Nalebuff [15] posit a more general form of a utility function where χ corresponds to v in our model, α corresponds to (τ_c, τ_p) in our model, and z is related to our good 0.

Each shopkeeper k has access to an exchange technology that allows him to set up a trading post in any pair of goods $(\nu^k, \widehat{\nu}^k) \in (\mathcal{V} \cup \{0\})^2$, where we assume without loss of generality $\nu^k \leq \widehat{\nu}^k$. Shopkeepers cannot produce goods and trading posts are in separate locations. All shopkeepers have identical linear preferences over all consumable goods and incur a fixed disutility cost κ (per capita) each period that their shop is open. Our convention is to name a shopkeeper who does not open $(0,0)$, one that trades in money by $(0, \nu^k)$, and a barter post by $\left(\nu^k > 0, \widehat{\nu}^k > 0\right)$. Let $C^k(t)$ denote shopkeeper k's consumption of good ν^k, and $\widehat{C}^k(t)$ denote shopkeeper k's consumption of good $\widehat{\nu}^k$ in period t. Shopkeepers also discount the future at rate β. We assume that there is free entry by shopkeepers.

We assume that households are anonymous (their identity (h) and history of past actions are unobservable), while tastes and money holdings $(\theta, m(t))$ are observable. Our assumption about anonymity is consistent with shopkeepers keeping track of the distribution of household money holdings and tastes, but not specific identities as in Jovanovic and Rosenthal [32]. Shopkeeper identity, the varieties of goods at their shop, consumption, and cash $(k, (\nu^k, \widehat{\nu}^k), (C^k(t), \widehat{C}^k(t)), M^k(t))$ are publicly observable. Furthermore, there is no cross location communication during a period. Hence, consumption at one trading post cannot be conditioned on production at a different trading post during the period.

The timing in any period is as follows. At the beginning of the period, agents receive lump sum transfers. In stage 1 (which can possibly be an infinite number of rounds), shopkeepers choose whether to open a trading post (thereby incurring cost κ). At this stage, any trading post k specializing in goods $(\nu^k, \widehat{\nu}^k)$ announces a (possibly) taste and money contingent allocation of the quantity of variety ν^k (or $\widehat{\nu}^k$) good for variety $\widehat{\nu}^k$ (or ν^k) good. In stage 2, each household member directs his or her search to one (and only one) trading post or stays home.[5] Then production, exchange, and consumption decisions simultaneously take place at each post according to the announced allocation rule, after

[5]In this sense, the environment shares some similarity to the directed search model of Corbae, Temzelides, and Wright [19]. While that paper restricted matches to bilateral meetings between households, here we restrict matches to be between a household member and a shopkeeper where the shopkeeper can be in many such matches (which is why we call it multilateral matching) but the household cannot.

which workers and shoppers return home.[6]

2.3 Planner's Problem

In each period, the planner chooses where to locate trading posts in variety space $\{(\nu^k, \widehat{\nu}^k)\}_{(M;t)}$, which trading post(s) to direct the household (if at all) $\{(k_p, k_c, \nu_p, \nu_c)\}_{(\theta,m;t)}$, and the allocation of goods and money among households $\{(\ell, c, \delta_p, \delta_c)\}_{(\theta,m;t)}$ and shopkeepers at each post $\{(C^k, \widehat{C}^k, D^k)\}_{(M;t)}$.[7] In particular the planner chooses $\{(k_p, k_c, \nu_p, \nu_c, \ell, c, \delta_p, \delta_c)\}_{(\theta,m;t)}$, $\{(\nu^k, \widehat{\nu}^k, C^k, \widehat{C}^k, D^k)\}_{(M;t)}$ to maximize

$$\sum_t \sum_k \beta^t \begin{pmatrix} \int_{hc(k,\nu^k;t)\cup hc(k,\widehat{\nu}^k;t)} c(\theta, m; t) d\mu(\theta, m; t) \\ -\varepsilon \int_{hp(k,\nu^k;t)\cup hp(k,\widehat{\nu}^k;t)} \ell(\theta, m; t) d\mu(\theta, m; t) \end{pmatrix} \tag{2.1}$$

where $\mu(\theta, m; t)$ is the distribution of money holdings across households of type θ and money holdings m in period t, where

$$hc\left(k, \nu^k; t\right) = \left\{(\theta, m) : k_c(\theta, m; t) = k, \nu_c(\theta, m; t) = \nu^k\right\}$$

and

$$hp(k, \nu^k; t) = \left\{(\theta, m) : k_p(\theta, m; t) = k, \nu_p(\theta, m; t) = \nu^k\right\}$$

are the sets of households directed by the planner to consume ν^k and produce ν^k at trading post k, respectively.[8] Since by the assumptions laid out in the environment, the planner cannot direct a worker and/or a shopper to two different trading posts in any one period, we

[6] In an earlier version of this paper, we allowed for the possibility that shopkeepers would change their announced allocation rule when households actually arrived at their shop at stage 2. Such deviations are punishable and were non-profitable in our environment. In order to keep notation and analysis easy, we maintain this simple assumption.

[7] One can think of the planner making proposals which household members and shopkeepers can accept or reject (where rejection implies that the agent is precluded from exchange that period).

[8] In writing the objective function this way, we have assumed that if $|v_c(\theta, m; t) - \tau_c| > \frac{x}{2}$, then $c(\theta, m; t) = 0$. This is without loss of generality since such a household is indifferent to this allocation and the positive consumption could be reallocated to someone who values the consumption.

know that any two $hc\left(k,\nu^k;t\right)$ sets are disjoint as are any two $hp\left(k,\nu^k;t\right)$.[9] Note, in order to minimize notation, we assume that the planner does not direct a shopper who obtains no utility from consumption to a trading post.

This objective is maximized subject to the following constraints. First, resource feasibility at each trading post requires that for all t,

$$\begin{aligned}
C^k(M;t) \leq & \quad \int_{hp(k,\nu^k;t)} \ell(\theta,m;t)d\mu(\theta,m;t) - \int_{hc\left(k,\nu^k;t\right)} c(\theta,m;t)d\mu(\theta,m;t) \quad (2.2)\\
\widehat{C}^k(M;t) \leq & \quad \int_{hp\left(k,\widehat{\nu}^k;t\right)} \ell(\theta,m;t)d\mu(\theta,m;t) - \int_{hc\left(k,\widehat{\nu}^k;t\right)} c(\theta,m;t)d\mu(\theta,m;t) \quad (2.3)\\
D^k(M;t) \leq & \qquad \int_{hc\left(k,\nu^k;t\right)\cup hc\left(t,k,\widehat{\nu}^k\right)} \delta_c(\theta,m;t)d\mu(\theta,m;t)\\
& \qquad - \int_{hp\left(k,\nu^k;t\right)\cup hp\left(k,\widehat{\nu}^k;t\right)} \delta_p(\theta,m;t)d\mu(\theta,m;t) \qquad (2.4)
\end{aligned}$$

where $M^k(t+1) = M^k(t) + D^k(t) \geq 0$. Second, shopkeeper participation requires that for all t,

$$\sum_{\eta=t}^{\infty} \beta^{\eta-t} \iota_{\left\{(\nu^k(M;\eta),\widehat{\nu}^k(M;\eta))\neq(0,0)\right\}} \kappa \leq \sum_{\eta=t}^{\infty} \beta^{\eta-t}\left(C^k\left(M;\eta\right) + \widehat{C}^k\left(M;\eta\right)\right) \qquad (2.5)$$

where $\iota_{\{\cdot\}}$ is an indicator function which takes on the value 1 if the argument $\{\cdot\}$ is true (i.e. when the shop is open). Third, household participation at two different shops requires that if $k_p(\theta,m;t) \neq k_c(\theta,m;t)$ or $\hat{\nu}^k(M;t) = 0$, then for all t,

$$\begin{aligned}
\nu(\theta,m;t) &= c(\theta,m;t) - \epsilon\ell(\theta,m;t) + \beta\nu\left(\theta, m+\eta(t+1)+\delta_p(\theta,m;t)-\delta_c(\theta,m;t);t+1\right)\\
&\geq \max\left\{\begin{array}{c} c(\theta,m;t) + \beta\nu\left(\theta,m+\eta(t+1)-\delta_c(\theta,m;t);t+1\right),\\ -\epsilon\ell(\theta,m;t) + \beta\nu\left(\theta,m+\eta(t+1)+\delta_p(\theta,m;t);t+1\right)\\ \beta\nu\left(\theta,m+\eta(t+1);t+1\right)\end{array}\right\} \qquad (2.6)
\end{aligned}$$

[9] Specifically, for instance,

$$hc(k,v^k;t) \cap hc(k,\widehat{v}^k;t) = \varnothing.$$

where $\min\{\delta_c, 0\} - \min\{\delta_p, 0\} \leq m$. Household participation at a barter shop requires that if $k_p(\theta, m; t) = k_c(\theta, m; t) = k$ and $\nu^k(M; t) \neq 0$, then

$$\nu(\theta, m; t) = c(\theta, m; t) - \epsilon\ell(\theta, m; t) + \beta\nu\left(\theta, m + \eta(t+1); t+1\right) \geq \beta\nu\left(\theta, m + \eta(t+1); t+1\right) \tag{2.7}$$

which is equivalent to $c(\theta, m; t) - \epsilon\ell(\theta, m; t) \geq 0$.[10] Finally, consistency requires:

$$\overline{m}(t) = \sum_k M^k(t) + \int m d\mu\left(\theta, m; t\right).$$

We will show that the solution to the planner's problem ((2.1) subject to (2.2)-(2.7)) attains a simple upper bound under certain assumptions. We now provide some intuition for how we generate the result given the complicated nature of the planning problem. The potentially hard part of the problem, which is the heart of the paper, is to decide which shops to open $\{(\nu^k, \widehat{\nu}^k)\}$ and where to direct both shoppers $hc\left(k, \nu^k; t\right)$ and workers $hp\left(k, \nu^k; t\right)$.

The first key insight comes from manipulating the objective function using (2.2), (2.3), and (2.5). Specifically, substituting (2.2) and (2.3) into (2.5) evaluated at $t = 0$ yields

$$\sum_{t=0}^{\infty} \beta^t \iota_{\left\{(\nu^k(M;t), \widehat{\nu}^k(M;t)) \neq (0,0)\right\}} \kappa \leq \sum_{t=0}^{\infty} \beta^t \left(\begin{array}{c} \int_{hp(k,\nu^k;t) \cup hp(k,\widehat{\nu}^k;t)} \ell(\theta, m; t) d\mu(\theta, m; t) \\ - \int_{hc(k,\nu^k;t) \cup hc(k,\widehat{\nu}^k;t)} c(\theta, m; t) d\mu(\theta, m; t) \end{array} \right).$$

Since the shopkeeper's utility doesn't enter the planner's objective, the above shopkeeper's participation constraint will bind, so that substituting for

$$\int_{hc(k,\nu^k;t) \cup hc(k,\widehat{\nu}^k;t)} c(\theta, m; t) d\mu(\theta, m; t)$$

[10]Note that if a household is directed to a barter shop then $\delta_p = \delta_c = 0$.

into the objective function (2.1) yields

$$\sum_t \beta^t \left[(1-\varepsilon)\int_{\cup_k \left(hp(k,\nu^k;t)\cup hp(k,\widehat{\nu}^k;t)\right)} \ell(\theta,m;t)d\mu(\theta,m;t) - \sum_k \iota_{\left\{(\nu^k(M;t),\widehat{\nu}^k(M;t))\neq(0,0)\right\}}\kappa\right] \tag{2.8}$$

using the fact that

$$\sum_k \int_{hp(k,\nu^k;t)\cup hp(k,\widehat{\nu}^k;t)} \ell(\theta,m;t)d\mu(\theta,m;t) = \int_{\cup_k \left(hp(k,\nu^k;t)\cup hp(k,\widehat{\nu}^k;t)\right)} \ell(\theta,m;t)d\mu(\theta,m;t).$$

Since the utility function is linear, the distribution of consumption does not affect the objective function and the planner's objective is to direct workers to open shops (i.e. $\cup_k \left(hp(k,\nu^k;t)\cup hp(k,\widehat{\nu}^k;t)\right)$) to maximize economywide output net of the costs of opening trading posts.[11]

The second key insight is to simplify the complications associated with directing workers to shops (i.e. $\cup_k \left(hp(k,\nu^k;t)\cup hp(k,\widehat{\nu}^k;t)\right)$). Instead of focusing on which worker is directed to a given shop, we focus on the set of workers for whom exchange at that shop is feasible. Let

$$S(\nu,\widetilde{\nu}) = \begin{cases} \left\{(\tau_c,\tau_p)\in\left[\nu-\frac{x}{2},\nu+\frac{x}{2}\right]\times\left[\widetilde{\nu}-\frac{x}{2},\widetilde{\nu}+\frac{x}{2}\right]\right\} & \text{if } \nu\neq 0 \\ \left\{(\tau_c,\tau_p)\in(0,1]\times\left[\widetilde{\nu}-\frac{x}{2},\widetilde{\nu}+\frac{x}{2}\right]\right\} & \text{if } \nu=0 \\ \varnothing & \text{if } \widetilde{\nu}=0 \end{cases} \tag{2.9}$$

for $(\nu,\widetilde{\nu})\in(\mathcal{V}\cup\{0\})^2$ denote an "exchange neighborhood" which obeys the mod 1 arithmetic.[12] An exchange neighborhood represents the set of households who could produce good $\widetilde{\nu}$ in exchange for goodν. As can be seen from this definition, an exchange neighborhood with $\nu\neq 0$ requires a double coincidence of wants while an exchange neighborhood with $\nu=0$ does not. For a barter shop k that trades $\left(\nu^k(t),\widehat{\nu}^k(t)\right)\in\mathcal{V}^2$,

[11] Note that this is not to say that the distribution of consumption is unimportant for the problem. Obviously, this will affect the household participation constraints (2.6)-(2.7).

[12] For example,

$$S(0,\frac{x}{4}) = (0,1]\times\left(\left(0,\frac{3x}{4}\right]\cup\left[1-\frac{x}{4},1\right]\right).$$

we know that if a worker is producing $\widehat{\nu}^k$, then by definition $\tau_p \in \left[\widehat{\nu}^k - \frac{x}{2}, \widehat{\nu}^k + \frac{x}{2}\right]$. In the case of a barter shop, the household participation constraint (2.7) implies the shopper from that household should consume good ν^k (or else $c(\theta, m; t) - \epsilon\ell(\theta, m; t) \geq 0$ is violated); hence $\tau_c \in \left[\nu^k - \frac{x}{2}, \nu^k + \frac{x}{2}\right]$. Thus $hp(k, \widehat{\nu}^k; t) \subset S(\ \nu^k(t), \widehat{\nu}^k(t))$. For a monetary shop k that trades $\left(\nu^k(t), \widehat{\nu}^k(t)\right) \in \{0\} \times \mathcal{V}$, we know $hp(k, 0; t) = \varnothing$ and $hp(k, \widehat{\nu}^k; t) \subset S(0, \widehat{\nu}^k(t))$ since workers cannot produce money and if a worker is producing $\widehat{\nu}^k$, then $\tau_p \in \left[\widehat{\nu}^k - \frac{x}{2}, \widehat{\nu}^k + \frac{x}{2}\right]$ by definition. Thus,

$$\cup_k \left(hp(k, \nu^k; t) \cup hp(k, \widehat{\nu}^k; t)\right) \subset \cup_k \left(S(\nu^k(t), \widehat{\nu}^k(t)) \cup S(\widehat{\nu}^k(t), \nu^k(t))\right).$$

In this case,

$$\begin{aligned} & \sum_t \beta^t \left[(1-\varepsilon) \int_{\cup_k \left(hp(k,\nu^k;t) \cup hp(k,\widehat{\nu}^k;t)\right)} \ell(\theta, m; t) d\mu(\theta, m; t) - \sum_k \iota_{\left\{(\nu^k(M;t), \widehat{\nu}^k(M;t)) \neq (0,0)\right\}} \kappa\right] \\ \leq \ & \sum_t \beta^t \left[(1-\varepsilon) \int_{\cup_k \left(S(\nu^k(t),\widehat{\nu}^k(t)) \cup S(\widehat{\nu}^k(t),\nu^k(t))\right)} \ell(\theta, m; t) d\mu(\theta, m; t) - n(t)\kappa\right] \end{aligned} \tag{2.10}$$

where $n(t) = \sum_k \iota_{\left\{(\nu^k(M;t), \widehat{\nu}^k(M;t)) \neq (0,0)\right\}}$ denotes the number of trading posts opened at time t. Note that by using exchange neighborhoods, we obtain an upper bound of the objective (2.8) that is free of the hp decision rule. What we will show below (in Theorem 1) is that the solution of (2.1) subject to (2.2)-(2.7) actually attains this upper bound (2.10) under certain conditions.

The third key insight comes from understanding how opening trading posts covers the type space in terms of exchange neighborhoods. Specifically, a planner who proposes that shopkeeper k opens a barter trading post offering goods of variety $(\nu^k, \widehat{\nu}^k) \in \mathcal{V}^2$ at time t affords households with two exchange neighborhoods. For example, assume $x \in \left(\frac{1}{3}, \frac{1}{2}\right)$. Figure 1a represents the exchange neighborhoods associated with one barter trading post offering goods of variety $\left(\frac{1}{4}, \frac{3}{4}\right)$. In particular, since with barter both the worker and shopper must go to the same trading post in order to satisfy the household participation con-

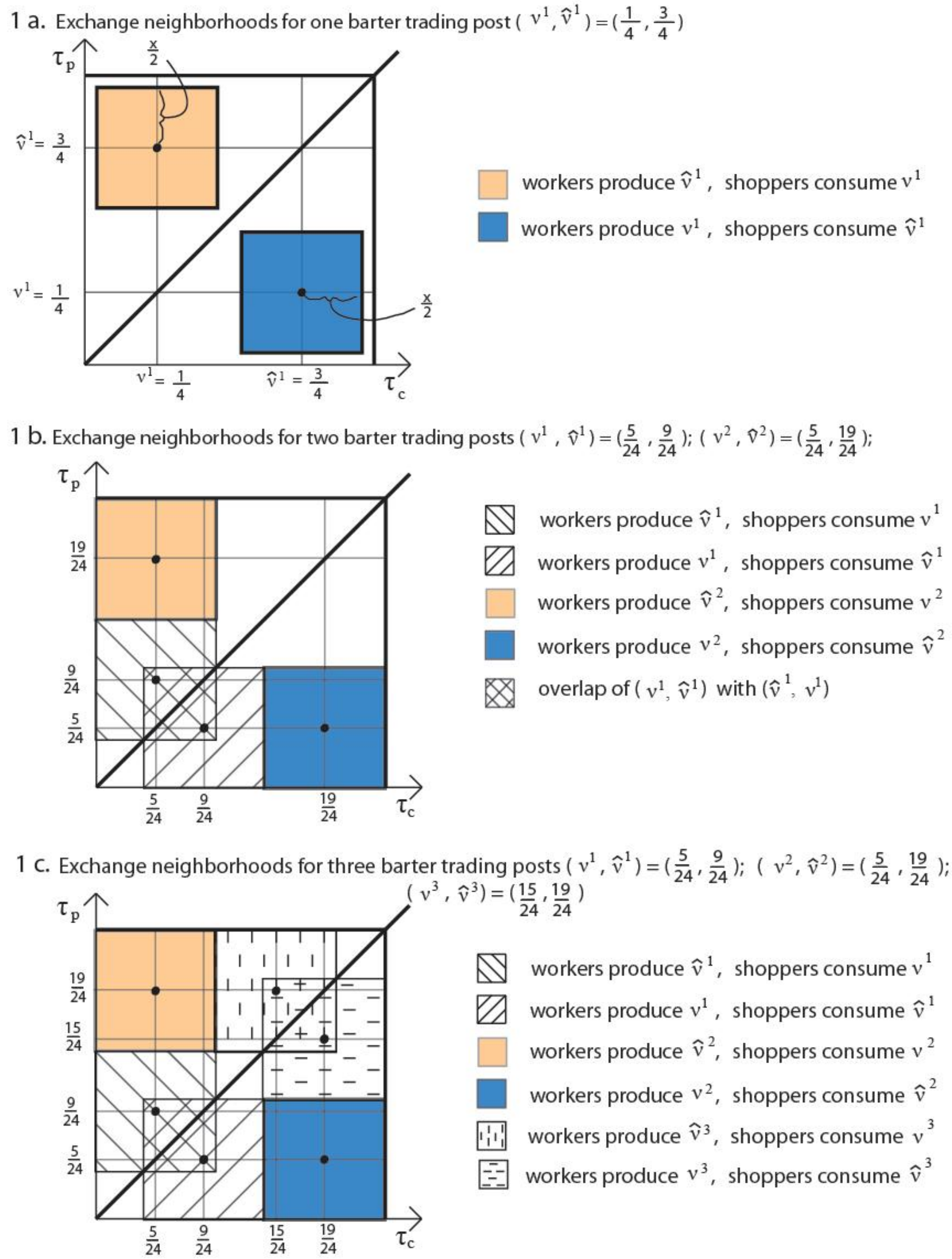


Figure 2.1: Barter Exchange Neighborhoods

straint (2.7), the shopkeeper can facilitate exchange between households with $(\tau_c, \tau_p) \in \left[\frac{1}{4} - \frac{x}{2}, \frac{1}{4} + \frac{x}{2}\right] \times \left[\frac{3}{4} - \frac{x}{2}, \frac{3}{4} + \frac{x}{2}\right]$ who come to his shop to exchange their variety $\frac{3}{4}$ good for a variety $\frac{1}{4}$ good *and* households with $(\tau_c, \tau_p) \in \left[\frac{3}{4} - \frac{x}{2}, \frac{3}{4} + \frac{x}{2}\right] \times \left[\frac{1}{4} - \frac{x}{2}, \frac{1}{4} + \frac{x}{2}\right]$ who come to his shop to exchange their variety $\frac{1}{4}$ good for a variety $\frac{3}{4}$ good. By opening this first trading post, the planner covers two squares with an area of size $2x^2$ of the type space. Note that not everyone can produce (and hence consume) with this one shop which "partially covers" the taste space (i.e. $2x^2 < 1$ with $x \in \left(\frac{1}{3}, \frac{1}{2}\right)$). Figure 1b represents the exchange neighborhoods associated with two barter trading posts offering goods of variety $\left(\frac{5}{24}, \frac{15}{24}\right)$ and $\left(\frac{9}{24}, \frac{19}{24}\right)$. With two barter posts and $x > \frac{1}{3}$, there is no way for the planner to arrange the two posts without having some "overlap" of the two exchange neighborhoods (i.e. $S\left(\frac{5}{24}, \frac{15}{24}\right) \cap S\left(\frac{9}{24}, \frac{19}{24}\right) \neq \varnothing$). At the same time, since $x < \frac{1}{2}$, it is not possible to cover the entire taste space with the two shops. By opening the second trading post, the planner has generated four exchange neighborhoods; however, because of the overlap of these neighborhoods the coverage is less than $4x^2$. Given that it is costly to open shops, it is inefficient to have any overlap. Finally, Figure 1c represents the exchange neighborhoods associated with three barter trading posts offering goods of variety $\left(\frac{5}{24}, \frac{9}{24}\right)$, $\left(\frac{5}{24}, \frac{19}{24}\right)$, and $\left(\frac{15}{24}, \frac{19}{24}\right)$. While opening a third trading post generates six exchange neighborhoods, the marginal increase in coverage will be lower due to the nonlinear rise in overlap. In summary, opening n barter trading posts generates $2n$ exchange neighborhoods and covers $2nx^2$ minus the area of overlap (where overlap is increasing in n at a (weakly) increasing rate).[13]

Unlike the two square exchange neighborhoods associated with each barter post (as evident in the first line of (2.9)), the exchange neighborhood of a $\left(0, \hat{\nu}^k\right)$ monetary shop is an entire strip centered on the $\tau_p = \hat{\nu}^k$ line (as evident in the second line of (2.9)). In particular, Figure (2.2)a represents the exchange neighborhood associated with one money trading post offering goods of variety $\left(0, \frac{1}{4}\right)$. Specifically, workers from households with $\tau_p \in \left[\frac{1}{4} - \frac{x}{2}, \frac{1}{4} + \frac{x}{2}\right]$ are willing to exchange production of their variety $\frac{1}{4}$ good for money

[13]An upper bound on the number of barter shops necessary to cover the taste space is $\frac{\lceil \frac{1}{x} \rceil \lceil \frac{1}{x} + 1 \rceil}{2}$ where $\lceil z \rceil$ denotes the ceiling of z, which is the smallest integer greater than or equal to z.

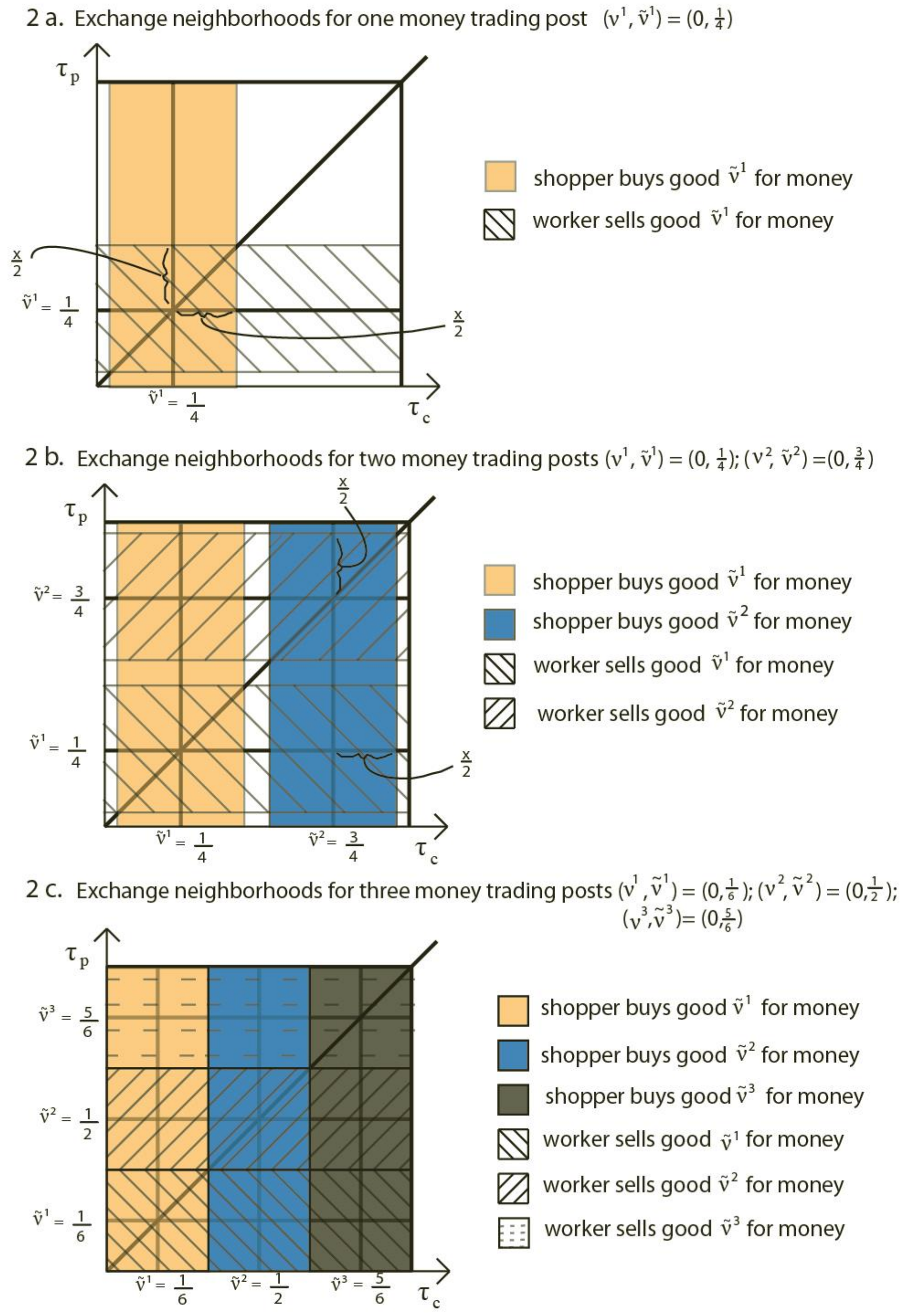


Figure 2.2: Monetary Exchange Neighborhoods

provided the shopper of the household will be directed to a money post in some future period to exchange its money for a variety $\hat{\nu}^{k'} \in \left[\tau_c - \frac{x}{2}, \tau_c + \frac{x}{2}\right]$ good such that the household participation constraint (2.6) is satisfied. Also pictured in Figure (2.2)a is the set (vertical strip) of shoppers who if they had money could go to the monetary shop to obtain good $\frac{1}{4}$ for their money. If in every future period the planner only opens the $\left(0, \frac{1}{4}\right)$ monetary shop, then the only households who could conceivably satisfy the participation constraint would be those for whom $(\tau_c, \tau_p) \in \left[\frac{1}{4} - \frac{x}{2}, \frac{1}{4} + \frac{x}{2}\right] \times \left[\frac{1}{4} - \frac{x}{2}, \frac{1}{4} + \frac{x}{2}\right]$ pictured in Figure (2.2)a (i.e. the intersection of the above two strips). However, a planner could always open monetary shops in other regions of the variety space in the future and keep others closed.[14] Obviously, the period in which the planner opens the future post (and the shopper gets to consume) will affect the participation constraint and hence β matters for this decision. Figure (2.2)b considers two monetary trading posts offering goods of variety $\left(0, \frac{1}{4}\right)$ and $\left(0, \frac{3}{4}\right)$. With $x \in \left(\frac{1}{3}, \frac{1}{2}\right)$, the two monetary exchange neighborhoods (and the associated vertical strips of shoppers who could go to those monetary posts) do not cover the entire type space but unlike the barter case, there is no overlap. Finally, Figure (2.2)c graphs three monetary trading posts $\left(0, \frac{1}{6}\right)$, $\left(0, \frac{1}{2}\right)$, and $\left(0, \frac{5}{6}\right)$. While there is overlap with $x \in \left(\frac{1}{3}, \frac{1}{2}\right)$, the three monetary exchange neighborhoods cover the entire type space (unlike the case for three barter posts).[15] If the same shops are open next period, then since there is full coverage of the type space, all shoppers have a place to spend their money (and there is no need for additional shops). It is simplest to see how money resolves the double coincidence problem in this case. Take two households of type $(\tau_c, \tau_p) = \left(\frac{5}{6}, \frac{1}{6}\right)$ and $(\tau_c, \tau_p) = \left(\frac{1}{6}, \frac{1}{2}\right)$. With $x \in \left(\frac{1}{3}, \frac{1}{2}\right)$, the $\left(\frac{5}{6}, \frac{1}{6}\right)$ household can produce for the $\left(\frac{1}{6}, \frac{1}{2}\right)$ household, but the reverse is not possible. However, the $\left(0, \frac{1}{6}\right)$ monetary shop facilitates trade between these two households. In summary, we can add up to $\lfloor \frac{1}{x} \rfloor$ monetary shops

[14]In fact, it is clear that if there were only one monetary trading post $\left(0, \hat{\nu}^k\right)$ open every period, the planner would never keep it in the same location in variety space since the associated $S(\hat{\nu}^k, \hat{\nu}^k) = \left[\hat{\nu}^k - \frac{x}{2}, \hat{\nu}^k + \frac{x}{2}\right] \times \left[\hat{\nu}^k - \frac{x}{2}, \hat{\nu}^k + \frac{x}{2}\right]$ square would be dominated by one barter trading post $(\nu^k, \hat{\nu}^k)$ which generates the two squares $S(\nu^k, \hat{\nu}^k)$ and $S(\hat{\nu}^k, v^k)$.

[15]For this latter assertion, see case 1 of the proof of Theorem 1.

without generating any overlap of the exchange neighborhoods and each shop covers an area of size x.[16] With $\lceil \frac{1}{x} \rceil$ monetary shops, the entire type space can be covered and there is no reason to open an additional shop and incur κ.[17] Moreover, if $\lfloor \frac{1}{x} \rfloor \neq \lceil \frac{1}{x} \rceil$, then the $\lceil \frac{1}{x} \rceil^{th}$ shop only adds an area of size $1 - \lfloor \frac{1}{x} \rfloor x < x$. Finally, in order for all workers in a given monetary exchange neighborhood $S(0, \widehat{\nu}^k)$ to produce, how many periods the shopper must wait before spending money is critical to satisfying the participation constraint (2.6).

Next we characterize the solution for the planner's problem (2.1) subject to (2.2)-(2.7). Specifically,

Theorem 11 *The solution for the planner's problem can be characterized as follows.*

Case 1. For $x < \frac{1}{2}$, if $\kappa \leq \left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon)$ and households are patient enough such that $\beta\left(1 - \lceil \frac{1}{x} \rceil \kappa\right) - \epsilon \geq 0$, then the unique solution to the planner's problem involves opening only $\lceil \frac{1}{x} \rceil$ monetary shops (Full Coverage).[18] *If $x(1-\epsilon) > \kappa > \left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon)$ and households are patient enough such that $\beta^2\left(1 - \frac{\kappa}{x}\right) - \epsilon \geq 0$, then the unique solution involves opening only $\lfloor \frac{1}{x} \rfloor$ monetary shops (Partial Coverage). If $\kappa \geq x(1-\epsilon)$, then no shop will be opened (No Coverage).*

Case 2. For $\frac{1}{2} \leq x < \frac{2}{3}$, if $\kappa < 2(1-x)^2(1-\epsilon)$ and households are patient enough such that $\beta(1-2\kappa) - \epsilon \geq 0$ then the unique solution involves opening two monetary shops and all workers produce every period (Full Coverage). If $2(1-x)^2(1-\epsilon) \leq \kappa \leq \left(1 - 2(1-x)^2\right)(1-\epsilon)$, then the unique solution involves opening a single barter shop (Partial Coverage). If $\left(1 - 2(1-x)^2\right)(1-\epsilon) > \kappa$, then no shop will be opened (No Coverage).

Case 3. For $x \geq \frac{2}{3}$, if $\kappa < 2(1-x)^2(1-\epsilon)$ then a solution to the planner's problem involves opening two barter shops. If households are patient enough such that $\beta(1-2\kappa) - \epsilon \geq 0$, then another solution involves opening two monetary shops. In either case all

[16] The floor $\lfloor z \rfloor$ denotes the largest integer less than or equal to z.

[17] Recall that $\lceil z \rceil$ denotes the ceiling of z, which is the smallest integer greater than or equal to z.

[18] By unique, we mean that the number of shops is uniquely determined. Which varieties are traded and who produces and consumes them is not unique. One source of this multiplicity is due to the mod arithmetic which ensures symmetry among households.

workers will produce (Full Coverage). If $2(1-x)^2(1-\epsilon) \leq \kappa \leq \left(1-2(1-x)^2\right)(1-\epsilon)$, *then the unique solution involves opening a single barter shop (Partial Coverage) and if* $\left(1-2(1-x)^2\right)(1-\epsilon) > \kappa$, *no shop will be opened (No Coverage).*

Proof. See appendix. ■

We break the characterization into cases depending on households' specialization of tastes (x) and the cost of opening shops (κ). Specifically, for each of three regions of x ranging from very selective to indiscriminate preferences, we then consider three possibilities for κ ranging from low to high costs. When x is small, strips associated with monetary neighborhoods cover more than two squares associated with barter neighborhoods. Furthermore, the placement of strips by the planner (i.e. location of monetary shops) can easily occur without overlap, but the placement of squares (i.e. location of barter shops) is prone to overlap (and figuring the least overlap way to place the shops takes many pages of proof). If κ is too high (i.e. $\kappa > x(1-\epsilon)$ so that the net utility associated with production by all workers in a monetary exchange neighborhood does not cover the shopkeeper's participation constraint), opening a shop is not feasible (no coverage). On the other hand, if κ is very low (i.e. $\left(1-\lfloor\frac{1}{x}\rfloor x\right)(1-\epsilon) > \kappa$ so that the net utility associated with the additional production by workers in the monetary exchange neighborhoods covers the cost of the last monetary shop added) and the sufficient conditions on β are met, the planner opens $\lceil\frac{1}{x}\rceil$ shops to fully cover the type space. The intermediate case, again provided the sufficient conditions on β are met, involves opening $\lfloor\frac{1}{x}\rfloor$ shops partially covering the type space without overlap of monetary exchange neighborhoods.

2.4 Examples

We illustrate several interesting corollaries associated with Theorem 1 via the next set of examples.

2.4.1 Monetary exchange can increase product variety

Here we provide an example that shows the introduction of money can increase product variety. Assume $\frac{1}{3} < x < \frac{11}{30}$. If

$$\left(2x^2 - (3x-1)^2\right)(1-\epsilon) < \kappa < (1-2x)(1-\epsilon) \tag{2.11}$$

and if money is not available, we show the optimal number of barter shops is 1 (so there are two varieties of goods in the economy).[19] Then we show that for the same parameter values, if money becomes available, the optimal number of monetary shops is 3 (so there are three varieties of goods in the economy).

To see this, we begin by noting that since $x < \frac{11}{30} < \frac{1}{2}$, the two exchange neighborhoods of a single barter shop can be located such that they do not overlap (See Figure (2.1)a). Hence the planner's objective function (2.10) with a single barter shop will take the value of $2x^2(1-\epsilon) - \kappa$. Since $x > \frac{1}{3}$ the exchange neighborhoods of two barter shops overlap (this is evident in Figure (2.1)b). In the appendix (Lemma 1) we show that the size of the overlap will be at least $(3x-1)^2$. Hence the planner's objective function (2.10) with two barter shops yields the value of $\left[4x^2 - (3x-1)^2\right](1-\varepsilon) - 2\kappa$. Now if the lower inequality in (2.11) holds, in the absence of money, the planner opens a single barter shop and each period two types of goods are produced in the economy.

With $\frac{1}{3} < x < \frac{11}{30} < \frac{1}{2}$, provided households are patient enough to satisfy the participation constraint (2.6), i.e. $1 - \frac{\kappa}{x} \geq \frac{\epsilon}{\beta^2}$ as given in case 1 of Theorem (1), two monetary shops deliver the value of $2x(1-\epsilon) - 2\kappa$ to the planner's objective function (2.10) under partial coverage (see Figure (2.2)b) and three monetary shops deliver the value $(1-\epsilon) - 3\kappa$ under full coverage (see Figure (2.2)c). If the upper inequality in (2.11) holds, then the planners opens three monetary shops and each period three types of goods are produced in the economy. This example illustrates the basic idea that the inefficiency

[19] If $x < \frac{11}{30}$ it follows $(1-2x) > 2x^2 - (3x-1)^2$.

associated with overlap of exchange neighborhoods when adding barter shops limits their adoption while adding money shops minimizes such inefficiencies.

2.4.2 The adoption of monetary exchange with lower costs of operating shops

In this example we show that as the cost of operating shops falls, the solution for the planner's problem changes from operating barter shops to opening monetary shops. Assume $\frac{1}{2} < x < \frac{2}{3}$. If the cost is in the interval

$$\kappa \in \left[2(1-x)^2(1-\epsilon), \left(1-2(1-x)^2\right)(1-\epsilon)\right],$$

the optimality of a single barter shop follows from Theorem 1, Case 2 (compare the areas covered in Figure (2.3)a with Figure (2.3)b). Specifically, the objective function (2.10) evaluated for one barter shop is given by $\left[1-2(1-x)^2\right](1-\varepsilon)-\kappa$ since the area covered by the barter exchange neighborhood in Figure (2.3)a is simply 1 minus the area not covered (the 2 squares in the northeast and southwest corners of Figure (2.3)a are of length $(1-x)^2$). Hence, provided $\kappa \leq \left[1-2(1-x)^2\right](1-\varepsilon)$, the planner opens at least the one barter shop (which would dominate one money shop since the upper bound of its objective (2.10) is given by $x(1-\varepsilon)-\kappa$). The condition $\kappa \geq 2(1-x)^2(1-\epsilon)$ guarantees that the cost is high enough that the additional value from covering the type space with two money shops versus the one barter shop is just associated with the area not covered above (i.e. $2(1-x)^2(1-\varepsilon)$); hence it is not worthwhile to open two money shops. But, if the cost of operating shops decreases (i.e. $\kappa < 2(1-x)^2(1-\epsilon)$), then provided households are sufficiently patient to satisfy the participation constraint (2.6), that is $\beta(1-2\kappa)-\varepsilon \geq 0$, the optimal solution for the planner's problem will be attained by opening two monetary shops covering the entire taste space (compare the areas covered in Figure (2.3)d with Figure (2.3)c). In this case, by decreasing κ we go from a barter economy where some households do not produce and do not consume, to a monetary economy where all households produce and consume.

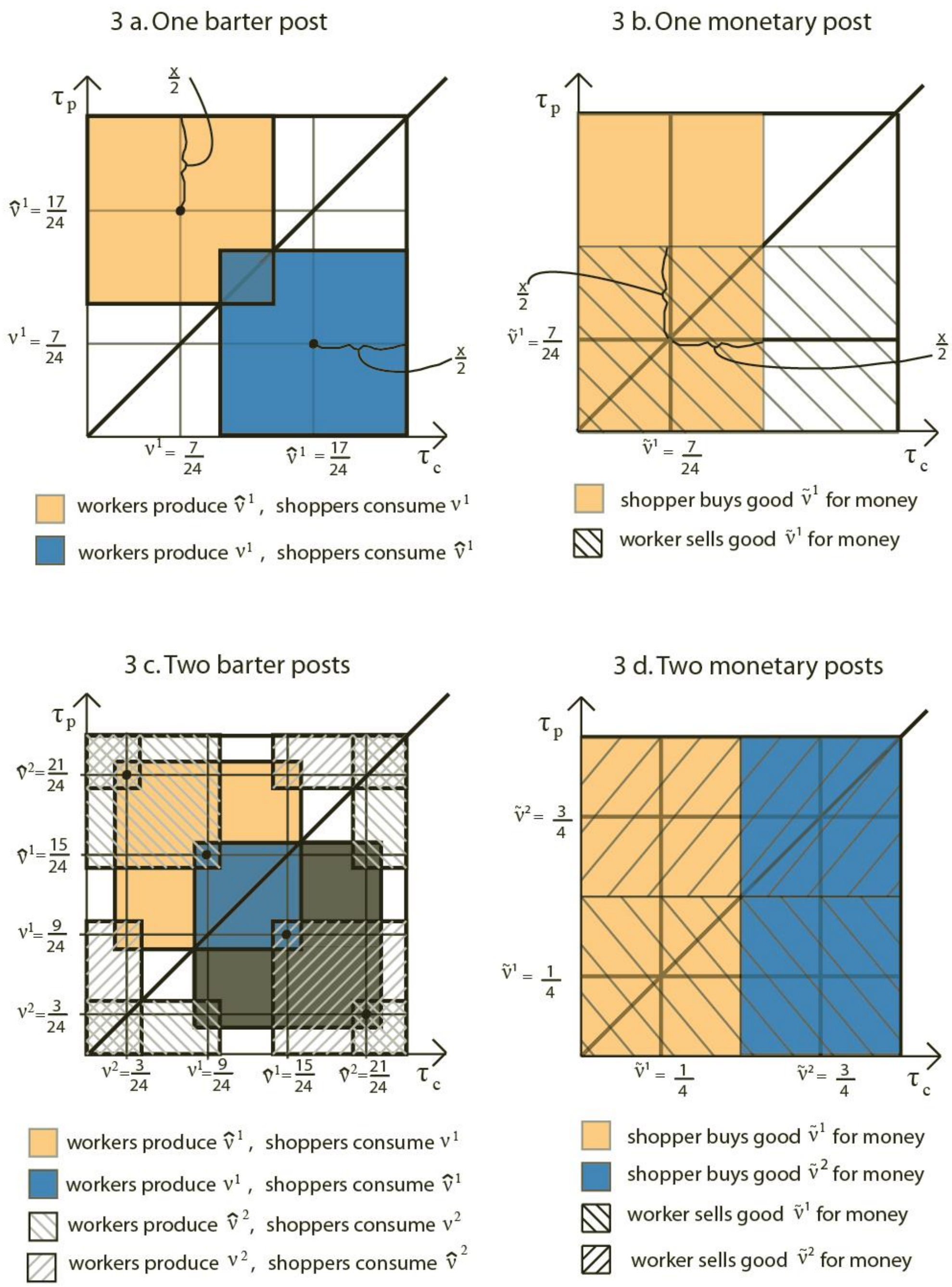


Figure 2.3: Barter to Monetary

2.4.3 Complementarity between monetary shops

Here we examine the positive feedback associated with opening monetary shops. In regions of partial coverage, we show that adding a barter shop generates negative feedback due to overlap, but adding a monetary shop generates positive feedback since it makes it easier to satisfy the household participation constraint. First we show that while a single barter shop may be incentive feasible, adding a second barter shop tightens the participation constraint for the existing shop due to overlap of the barter exchange neighborhoods. Then we show that while a single monetary shop is not incentive feasible since it is not possible for the planner to move the one shop around the variety space over time to satisfy the household participation constraint (in particular it would take at least two periods before consuming), adding a second monetary shop that covers the taste space (so that it only takes one period between producing and consuming) loosens the household participation constraint.

Assume $x = \frac{1}{2}$. Furthermore, assume

$$\beta^2 < \epsilon < \beta < \frac{1}{3}, \tag{2.12}$$

$$\frac{3}{4}(1-\epsilon) < 2\kappa, \tag{2.13}$$

$$\beta(1-2\kappa) - \epsilon > 0. \tag{2.14}$$

It is possible to show that the set of parameters satisfying (2.12)-(2.14) is not empty.[20]

At a single barter shop, trading $\left(\frac{1}{4}, \frac{3}{4}\right)$, half of the households produce and consume. From the $\frac{1}{2}$ units of goods produced, κ units is consumed by the shopkeeper in order to satisfy the shopkeeper participation constraint (2.5). The remaining $\frac{1}{2} - \kappa$ units are consumed

[20]Notice for $\beta < \frac{1}{3}$, we have

$$\beta^2 < \frac{1}{\frac{4}{\beta} - 3} < \beta.$$

Now if ϵ is such that

$$\beta^2 < \epsilon < \frac{1}{\frac{4}{\beta} - 3},$$

satisfying (2.12), then the second inequality implies $\frac{\epsilon}{\beta} < 1 - \frac{3}{4}(1-\epsilon)$. For a κ satisfying $\frac{\epsilon}{\beta} < 1 - 2\kappa < 1 - \frac{3}{4}(1-\epsilon)$, the first inequality implies (2.14) and (2.13) follows from the second one.

by the $\frac{1}{2}$ shoppers visiting the shop, each consuming $1-2\kappa$. The household participation constraint (2.7), $(1-2\kappa)-\varepsilon>0$, is satisfied given (2.14). Due to overlap of their exchange neighborhoods, two barter shops cover at most $4x^2-(3x-1)^2=\frac{3}{4}$ measure of the taste space. By (2.13) we know that after paying the least necessary consumption (i.e. ε) to households in order to satisfy their participation constraint (2.7), since there are only $\frac{3}{4}$ units of goods remaining, there will not be enough to satisfy the shopkeeper participation constraint (2.5). In summary, given (2.13) while it is incentive feasible to set up one barter shop, opening two of them is incentive infeasible.

A single money shop provides production opportunities to $x=\frac{1}{2}$ measure of workers (the area of the strip). The shoppers of $x^2=\frac{1}{4}$ measure of these households can consume next period, while the shoppers of $x^2=\frac{1}{4}$ measure of them have to wait for at least two periods before consumption. Denote the set of households whose worker produces in period t and whose shopper has to wait for at least two periods before consumption by A_t. Let $\#(A_t)$ denote the measure of them, in which case $\#(A_t)\leq x=\frac{1}{2}$. Moreover, let $\bar{A}$ denote the supremum of $\{\#(A_t)\}_{t=0}^{\infty}$. Suppose the participation constraint can be satisfied for all households (we will derive a contradiction to this). For those who wait for at least two periods, the participation constraint (2.6) implies the continuation value on the equilibrium path after two periods$\nu(\theta,m;t+2)$ for $\theta\in A_t$ should exceed $\frac{\epsilon}{\beta^2}$ for each of them. Then

$$\#(A_t)\frac{\varepsilon}{\beta^2}\leq\int_{A_t\times\mathbb{R}_+}\nu(\theta,m;t+2)d\mu\left(\theta,m;t\right)\leq\int_{\Theta\times\mathbb{R}_+}\nu(\theta,m;t+2)d\mu\left(\theta,m;t\right)$$

where the first inequality follows from aggregating household participation constraints and the second inequality follows since the continuation value cannot be negative for *any* household. Every period at most $\frac{1}{4}+\bar{A}$ of workers produce, therefore (2.10) implies

$$\int_{\Theta\times\mathbb{R}_+}\nu(\theta,m;t+2)d\mu\left(\theta,m;t\right)\leq\frac{1}{1-\beta}\left[\left(\frac{1}{4}+\bar{A}\right)(1-\epsilon)-\kappa\right].$$

Since $\frac{\varepsilon}{\beta^2} > 1$, by definition, there exists some period t in which $\#(A_t)\frac{\varepsilon}{\beta^2} > \bar{A}$. Therefore we should have

$$\bar{A} \leq \frac{1}{1-\beta}\left[\left(\frac{1}{4} + \bar{A}\right)(1-\epsilon) - \kappa\right].$$

Since $\varepsilon > \beta^2$, $\beta < \frac{1}{3}$and $\kappa > \frac{3}{8}(1-\epsilon)$, the above inequality implies:

$$\bar{A} < \frac{1}{1-\beta}(\bar{A} - \frac{1}{8})(1-\beta^2) < \frac{4}{3}(\bar{A} - \frac{1}{8})$$

which in turn implies $\bar{A} > \frac{1}{2}$, contradicting $\#(A_t) \leq \frac{1}{2}$ for $\forall t$. On the other hand, two monetary shops cover the type space so that consumption and production can take place every period. Since κ must be provided to each of the shopkeepers to satisfy their participation constraint (2.5), the remaining output can be used to satisfy the household participation constraint (which follows from the assumption in (2.14)). In summary, given our assumptions (2.12)-(2.14), while it is not incentive feasible to open a single monetary shop, it is feasible to open two of them. The positive feedback to the participation constraint comes from the fact that the measure of the set of households who can produce one period and consume the next is $\frac{1}{4}$ with one monetary shop while it is a set of measure 1 with two monetary shops.

2.5 Implementation

In this section, we show that we can implement the solution to the planner's problem given in Theorem 1 as a subgame perfect equilibrium. Free entry will ensure that allocations yield zero profits for shopkeepers (who will not accumulate money). With full coverage, the endogenous cash-in-advance constraint yields a unit velocity quantity theory and households are treated identically. In particular, we consider implementing the planner's solution with linear prices rather than nonlinear prices.

Theorem 12 *In all cases of Theorem 1 where there is full coverage provided the money growth rate is small enough such that the household participation constraint (2.6) is sat-*

isfied (i.e. $\frac{\beta}{1+\gamma}\left(1-\lceil\frac{1}{x}\rceil\kappa\right)-\epsilon\geq 0$), then the unique solution to the planner's problem can be implemented as a subgame perfect equilibrium with $\lceil\frac{1}{x}\rceil$ monetary trading posts who each pay type independent wages $W(t)=\overline{m}(t)$ and charge type independent prices $P(t)=\frac{\overline{m}(t)}{1-\lceil\frac{1}{x}\rceil\kappa}$. Along the equilibrium path, shopkeepers earn zero profits and do not accumulate money. The equilibrium distribution of real money holdings is degenerate at $1-\lceil\frac{1}{x}\rceil\kappa$.

Proof. The public history of any shopkeeper k includes his stage 1 announced prices $P(t)$ and $W(t)$. Let $\mathcal{H}_t=0$ iff $\mathcal{H}_{t-1}=0$ (with initial condition $\mathcal{H}_{-1}=0$) and by the end of stage 1 of period t for any $i\in\left\{1,\ldots,\lceil\frac{1}{x}\rceil\right\}$, there is at least one monetary shop trading $\left(0,\frac{i-\frac{1}{2}}{\lceil\frac{1}{x}\rceil}\right)$, with price $P(t)=\frac{\overline{m}(t)}{1-\lceil\frac{1}{x}\rceil\kappa}$ and wage $W(t)=\overline{m}(t)$ and $\mathcal{H}_t=1$ otherwise.[21]

In any history $\mathcal{H}_t$, we consider the following strategy by a household of type $(\tau_c,\tau_p)\in\left(\frac{i-1}{\lceil\frac{1}{x}\rceil},\frac{i}{\lceil\frac{1}{x}\rceil}\right]\times\left(\frac{j-1}{\lceil\frac{1}{x}\rceil},\frac{j}{\lceil\frac{1}{x}\rceil}\right]$ for $i,j\in\left\{1,\ldots,\lceil\frac{1}{x}\rceil\right\}$. In stage two of any period t in history $\mathcal{H}_t=0$, the shopper of such a household visits the monetary shop which trades the pair $\left(0,\frac{i-\frac{1}{2}}{\lceil\frac{1}{x}\rceil}\right)$ with the lowest index k.[22] At that shop he spends all of his money holdings. The worker of such a household visits the monetary shop which trades the pair $\left(0,\frac{j-\frac{1}{2}}{\lceil\frac{1}{x}\rceil}\right)$ with the lowest index k'. At that shop she produces one unit of the good. If $\mathcal{H}_t=1$, both household members stay home.

We consider the following strategy by shopkeepers. In stage one of any period t in history $\mathcal{H}_{t-1}=0$, if no one has opened the i^{th} monetary shop trading $\left(0,\frac{i-\frac{1}{2}}{\lceil\frac{1}{x}\rceil}\right)$ with price $P(t)=\frac{\overline{m}(t)}{1-\lceil\frac{1}{x}\rceil\kappa}$ and wage $W(t)=\overline{m}(t)$, then some shopkeeper enters bearing the cost κ and announces such prices. Specifically, in the k^{th} round of the first stage, if none of the shopkeepers $z<k$ have opened the monetary shop to trade the pair $\left(0,\frac{mod_{\lceil\frac{1}{x}\rceil}(k)-\frac{1}{2}}{\lceil\frac{1}{x}\rceil}\right)$ at announced price $P(t)=\frac{\overline{m}(t)}{1-\lceil\frac{1}{x}\rceil\kappa}$ and wage $W(t)=\overline{m}(t)$, then the k^{th} shopkeeper incurs

[21] Note that $\mathcal{H}_t=0$ does not require all shopkeepers to follow the equilibrium strategy, just enough to cover the type space with equlibrium prices.

[22] By sending the household to the lowest index k when there are potentially multiple monetary shops offering good $\frac{i-\frac{1}{2}}{\lceil\frac{1}{x}\rceil}$ at the same prices, we are simply coordinating directed matching.

the cost κ and does so.[23] If a shopkeeper $z < k$ has opened the monetary shop at those prices, shopkeeper k does not open thereby avoiding the cost κ. In history $\mathcal{H}_{t-1} = 1$, all shopkeepers do not open.

To see there is not a profitable deviation by a shopkeeper in history $\mathcal{H}_{t-1} = 0$, notice that at the announced prices the shopkeeper receives zero profit net of the cost κ and accumulates no money. If a shopkeeper deviates and announces different prices/wages, then according to the above household strategy, no shopper or worker will visit their post leaving the shop to bear the cost κ without consuming anything or accumulating money. To see there is not a profitable deviation by a shopkeeper in history $\mathcal{H}_{t-1} = 1$, note that autarky is a subgame perfect equilibrium.

To see there is not a profitable deviation by a household in history $\mathcal{H}_t = 0$, first note that given that no other household will produce or consume at a monetary shop which is not following the shopkeeper strategy along the equilibrium path, it is not individually rational for any household to produce or consume at a deviant shop. In particular, given that shopkeepers do not start with money balances when $\mathcal{H}_t = 0$, then a deviant shopkeeper cannot offer higher wages or support lower prices because households will not visit his shop (as required by the household strategy). Hence, even if a shop announces lower prices or higher wages, the household strategy does not send the shopper or worker there. In stage two of any period t in history $\mathcal{H}_t = 0$, the continuation value of a type θ household with money holding m in period t using the above strategies is given by

$$\nu(\theta, m; t) = \frac{m}{\overline{m}(t)}\left(1 - \lceil\frac{1}{x}\rceil\kappa\right) - \varepsilon + \frac{\beta}{1-\beta}\left[\left(1 - \lceil\frac{1}{x}\rceil\kappa\right) - \varepsilon\right].$$

This follows from the household strategy that they spend all their money and the linearity of announced prices by shopkeepers along the equilibrium path. Given this, it is easy to verify

[23]This is only one algorithm to select potential shopkeepers to open trading posts, but there are many other algorithms to do the same thing. The algorithm simply assigns opening monetary shop $\left(0, \frac{i-\frac{1}{2}}{\lceil\frac{1}{x}\rceil}\right)$ to the shopkeepers with index k where $mod_{\lceil\frac{1}{x}\rceil}(k) = i$. It ensures that if a shopkeeper doesn't follow the ascribed strategy, another shopkeeper down the line will eventually enter with the ascribed strategy.

that the household participation constraint (2.6) holds provided $\frac{\beta}{1+\gamma}\left(1-\lceil\frac{1}{x}\rceil\kappa\right)-\varepsilon\geq 0$. That the household spends all of its money follows from linearity of the utility function, discounting, and possible money growth. In a history where $\mathcal{H}_t=1$, autarky is subgame perfect. ■

Notice that given this implementation uses linear prices, increasing the money growth rate weakens the household participation constraint thereby potentially eliminating this good equilibrium. While the consumption and production allocation is independent of the money growth rate, existence of the equilibrium itself is not superneutral. If instead of linear prices we had considered implementing the full coverage planner's solution with nonlinear prices, the household's participation constraint is independent of the money growth rate.[24]

In general, however, when there is partial coverage, prices need to be type dependent and nonlinear. The nonlinear type dependent prices can be inferred from the allocation described in table A.1 and A.2 of the proof of Theorem 1. Specifically, wages for all types are given by (2.19):

$$\begin{aligned} W(t) &= \lfloor\frac{1}{x}\rfloor xW(t-1)+\left(1-\lfloor\frac{1}{x}\rfloor x\right)W(t-2) \\ &\quad+\left(\frac{1}{\lfloor\frac{1}{x}\rfloor x}-2\left(1-\lfloor\frac{1}{x}\rfloor x\right)\right)\eta(t)+2\left(1-\lfloor\frac{1}{x}\rfloor x\right)\eta(t-1) \end{aligned}$$

but in each period only $\lfloor\frac{1}{x}\rfloor x$ of workers can produce while $1-\lfloor\frac{1}{x}\rfloor x$ of workers produce in alternating periods. As an example of the nonlinear goods prices, for $\theta=(\tau_c,\tau_p)\in\left[\frac{i}{\lfloor\frac{1}{x}\rfloor}-x,\frac{i-1}{\lfloor\frac{1}{x}\rfloor}+x\right]\times\left(\frac{j-1}{\lfloor\frac{1}{x}\rfloor},\frac{j}{\lfloor\frac{1}{x}\rfloor}-x\right)$ during an even period t^e, $P\left(\theta;t^e\right)=\infty$ and during

[24]Specifically, the nonlinear prices are given by

$$P(t)=\begin{cases}\frac{\overline{m}(t)}{1-\lceil\frac{1}{x}\rceil\kappa} & \text{if } m(t)\geq\overline{m}(t)\\ \infty & \text{otherwise}\end{cases}$$

which amounts to a two part tariff where the shopper pays $\overline{m}(t)$ to enter the trading post and receive $1-\lceil\frac{1}{x}\rceil\kappa$ while additional consumption costs $\frac{\overline{m}(t)}{1-\lceil\frac{1}{x}\rceil\kappa}$ per unit.

an odd period t^o, $P\left(\theta;t^o\right) = \frac{W(t^o-1)+\eta(t^o)+\eta(t^o-1)}{1-\frac{\kappa}{x}}$. These prices induce the following allocation for the above type θ household: during an even period t^e, $\ell(\theta;t^e) = 1$ and $c(\theta;t^e) = 0$ while during an odd period t^o, $\ell(t^o) = 0$ and $c(t^o) = 1 - \frac{\kappa}{x}$. Similarly, for $\theta = (\tau_c, \tau_p) \in \left(\frac{i-1}{\lfloor\frac{1}{x}\rfloor}, \frac{i}{\lfloor\frac{1}{x}\rfloor} - x\right) \times \left[\frac{j}{\lfloor\frac{1}{x}\rfloor} - x, \frac{j-1}{\lfloor\frac{1}{x}\rfloor} + x\right]$ during an even period t^e, $P\left(\theta;t^e\right) = \frac{W(t^e-1)+W(t^e-2)+\eta(t^e)+\eta(t^e-1)}{2\left(1-\frac{\kappa}{x}\right)}$ and during an odd period t^o, $P\left(\theta;t^o\right) = \infty$. These prices induce the following allocation for the above type θ household: during an even period t^e, $\ell(\theta;t^e) = 1$ and $c(\theta;t^e) = 2\left(1 - \frac{\kappa}{x}\right)$ while during an odd period t^o, $\ell(\theta;t^o) = 1$ and $c(\theta;t^o) = 0$. Thus, it is evident that for the case with partial coverage there is cross sectional variation in household consumption and production across time.

It is possible that we can implement an allocation that delivers the same aggregate utility as the planners solution (i.e. achieving the same value for the objective (2.1)) using type independent linear prices provided that γ is low enough and β is high enough (i.e. higher than that required with nonlinear prices to satisfy the household participation constraint). In this case the type independent wage rate is governed by the following difference equation

$$\begin{aligned} W(t) &= \lfloor\frac{1}{x}\rfloor x W(t-1) + \left(1 - \lfloor\frac{1}{x}\rfloor x\right) W(t-2) \\ &\quad + \eta(t) + \left(\frac{1}{\lfloor\frac{1}{x}\rfloor x} - 1\right) \eta(t-1). \end{aligned}$$

The type independent price is given by

$$P\left(t\right) = \frac{W\left(t\right)}{1 - \frac{\kappa}{x}}.$$

However, even with linear prices, the resulting allocation will exhibit cross sectional variation across time.

2.5.1 Example: Elimination of Pareto Dominated Equilibrium by Money Growth

From Theorem 1 it follows that for $\left(1 - x\lfloor\frac{1}{x}\rfloor\right)(1-\epsilon) > \kappa$ the optimal solution for the planner's problem (2.1) is attained by opening $\lceil\frac{1}{x}\rceil$ monetary shops to cover the type space, provided households are patient enough so their participation constraint is satisfied. In Theorem 2, we showed that this solution can be implemented with linear prices with zero money growth (i.e. $\beta\left(1 - \lceil\frac{1}{x}\rceil\kappa\right) - \epsilon \geq 0$). However, for a given set of beliefs, there might be other equilibria. In particular, if $\beta\left(1 - \frac{\kappa}{x^2\lfloor\frac{1}{x}\rfloor}\right) - \epsilon > 0$ then with a constant supply of money, it is possible to endow households with beliefs such that there exists a partial coverage equilibrium where only $\lfloor\frac{1}{x}\rfloor$ monetary shops open and trade the same goods each period for which $\left(x\lfloor\frac{1}{x}\rfloor\right)^2$ of the workers produce. The varieties of goods traded for money in these shops can be given by $(i-\frac{1}{2})x$ for $i \in \left\{1, 2, \ldots, \lfloor\frac{1}{x}\rfloor\right\}$ and households believe that no other shops will be opened. Notice that since the variety of goods traded do not change, although $x\lfloor\frac{1}{x}\rfloor$ workers have production opportunities, only $\left(x\lfloor\frac{1}{x}\rfloor\right)^2$ workers produce since only their shoppers have consumption opportunities.[25] Condition $\beta\left(1 - \frac{\kappa}{x^2\lfloor\frac{1}{x}\rfloor}\right) - \epsilon > 0$ guarantees the household participation constraint holds for them when the money supply is constant.

In order to see why more monetary shops don't open, notice that if all households believe that the only shops visited by workers and shoppers will be the $\lfloor\frac{1}{x}\rfloor$ monetary shops that partially cover the type space, then it is individually rational for each household member to stay home since along the equilibrium path there will be no money at the shops and no goods will be produced or consumed there.[26] With $\beta\left(1 - \frac{\kappa}{x^2\lfloor\frac{1}{x}\rfloor}\right) - \epsilon > 0$, the participation constraint for all workers producing at the existing shops holds, hence it is individually

[25]Actually, there are households with type $(\tau_c, \tau_p) \in \left(1 - x\left\lfloor\frac{1}{x}\right\rfloor, 1\right] \times \left(0, x\left\lfloor\frac{1}{x}\right\rfloor\right]$ who have production opportunities since their production speciality lies inside $\cup_i S(0, \left(i - \frac{1}{2}\right)x)$, but there are no shops under the assumed beliefs at which they can consume in the current or any future period.

[26]Actually, there are households with type $(\tau_c, \tau_p) \in \left(0, x\left\lfloor\frac{1}{x}\right\rfloor\right] \times \left(1 - x\left\lfloor\frac{1}{x}\right\rfloor, 1\right]$ who spend their initial endowment of money on consumption goods at one of the $\left\lfloor\frac{1}{x}\right\rfloor$ shops at time zero, but since their production speciality lies outside $\cup_i S(0, \left(i - \frac{1}{2}\right)x)$, these households do not work and receive wages in the future.

rational to produce and consume only at those $\lfloor \frac{1}{x} \rfloor$ monetary shops. In that case, it is not optimal for the shopkeepers to bear a cost in stage 1 and open another shop.

This partial coverage equilibrium is pareto dominated by the full coverage equilibrium provided

$$\left(\frac{1}{x}\right)^2 > \lfloor \frac{1}{x} \rfloor \lceil \frac{1}{x} \rceil. \tag{2.15}$$

Notice (2.15) implies consumption by households whose worker produces in the partial coverage case (i.e. $1 - \frac{\kappa}{x^2 \lfloor \frac{1}{x} \rfloor}$) is less than consumption by all households in the full coverage case (i.e. $1 - \lceil \frac{1}{x} \rceil \kappa$).

If there is positive money growth, the participation constraint for the partial coverage equilibrium described may not be satisfied (i.e. $\frac{\beta}{1+\gamma}\left(1 - \frac{\kappa}{x^2 \lfloor \frac{1}{x} \rfloor}\right) - \epsilon < 0$).[27] However, as long as $\frac{\beta}{1+\gamma}\left(1 - \lceil \frac{1}{x} \rceil \kappa\right) - \epsilon > 0$, Theorem 2 states that the full coverage solution can be implemented with linear prices. In summary, provided (2.15),

$$\beta\left(1 - \lceil \frac{1}{x} \rceil \kappa\right) - \epsilon > \beta\left(1 - \frac{\kappa}{x^2 \lfloor \frac{1}{x} \rfloor}\right) - \epsilon > 0$$

which guarantees that with zero money growth and the appropriate beliefs, both full and

[27]Notice that households with type $(\tau_c, \tau_p) \in \left(0, x\lfloor \frac{1}{x} \rfloor\right] \times \left(0, x\lfloor \frac{1}{x} \rfloor\right]$ want to use both their earned wage from last period and their money transfer in the current period to purchase goods, while households with type $(\tau_c, \tau_p) \in \left(0, x\lfloor \frac{1}{x} \rfloor\right] \times \left(x\lfloor \frac{1}{x} \rfloor, 1\right]$ just use their money transfers to consume since they do not produce. The period t price and wage are given by

$$P(t) = \frac{\overline{m}(t)\, x \lfloor \frac{1}{x} \rfloor}{\left(x \lfloor \frac{1}{x} \rfloor\right)^2 - \lfloor \frac{1}{x} \rfloor \kappa}$$

and

$$W(t) = \frac{\overline{m}(t)\, x \lfloor \frac{1}{x} \rfloor}{\left(x \lfloor \frac{1}{x} \rfloor\right)^2}.$$

The participation constraint for a worker in period t, can be written as $\beta \frac{W(t)}{P(t+1)} - \varepsilon > 0$, which is equivalent to $\frac{\beta}{1+\gamma}\left(1 - \frac{\kappa}{x^2 \lfloor \frac{1}{x} \rfloor}\right) - \varepsilon > 0$.

partial coverage are equilibria with linear prices and

$$\frac{\beta}{1+\gamma}\left(1-\lceil\frac{1}{x}\rceil\kappa\right)-\epsilon>0>\frac{\beta}{1+\gamma}\left(1-\frac{\kappa}{x^2\left\lfloor\frac{1}{x}\right\rfloor}\right)-\epsilon$$

which guarantees that the pareto dominated partial coverage equilibrium is eliminated with positive money growth while the full coverage equilibrium is not. It is simple to find regions of the parameter space (ϵ, κ, β, x, and γ) such that these constraints are satisfied. For instance, $\frac{1}{x}$ should be closer to $\lceil\frac{1}{x}\rceil$ than $\lfloor\frac{1}{x}\rfloor$ (so that $\left(\frac{1}{x}\right)^2>\lfloor\frac{1}{x}\rfloor\lceil\frac{1}{x}\rceil$). In short, a mild money growth eliminates the pareto dominated partial coverage equilibrium in favor of the optimal full coverage equilibrium.[28]

2.6 Conclusion

This paper is an attempt to integrate aspects of monetary theory familiar from Kiyotaki-Wright [35] (the idea of specialized tastes and production abilities) and Lagos-Wright [37] (the inclusion of multilateral matching in competitive markets with divisible money) with industrial organization along the lines of Hotelling [28]. An earlier version of this paper considered more general, smooth preferences over variety as in Kiyotaki and Wright [34] as well as concavity over the quantity of good consumed. These assumptions required nonlinear, type dependent prices and it was impossible to implement the planner's solution with linear prices unless we introduced private information. We solved the problem by considering the limit of this more general case.

One interesting extension is to consider how international payment systems influence the varieties of traded goods. Given that there is cross-sectional heterogeneity in regions of the parameter space, the model can generate cross-country variability in prices and output.

[28]This result is similar to a result for an overlapping generations economy in Cooper and Corbae [18], where a mild money growth eliminates the suboptimal equilibrium.

2.7 Appendix

Theorem 1. The solution for the planner's problem can be characterized as follows.

Case 1. For $x < \frac{1}{2}$, if $\kappa \leq \left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon)$ and households are patient enough such that $1 - \lceil \frac{1}{x} \rceil \kappa \geq \frac{\epsilon}{\beta}$, then the unique solution to the planner's problem involves opening only $\lceil \frac{1}{x} \rceil$ monetary shops (Full Coverage). If $x(1-\epsilon) > \kappa > \left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon)$ and households are patient enough such that $1 - \frac{\kappa}{x} \geq \frac{\epsilon}{\beta^2}$, then the unique solution involves opening only $\lfloor \frac{1}{x} \rfloor$ monetary shops (Partial Coverage). If $\kappa \geq x(1-\epsilon)$, then no shop will be opened (No Coverage).

Case 2. For $\frac{1}{2} \leq x < \frac{2}{3}$, if $\kappa < 2(1-x)^2(1-\epsilon)$ and households are patient enough such that $\beta(1-2\kappa) - \epsilon \geq 0$ then the unique solution involves opening two monetary shops and all workers produce every period (Full Coverage). If $2(1-x)^2(1-\epsilon) \leq \kappa \leq \left(1 - 2(1-x)^2\right)(1-\epsilon)$, then the unique solution involves opening a single barter shop (Partial Coverage). If $\left(1-2(1-x)^2\right)(1-\epsilon) > \kappa$, then no shop will be opened (No Coverage).

Case 3. For $x \geq \frac{2}{3}$, if $\kappa < 2(1-x)^2(1-\epsilon)$ then a solution to the planner's problem involves opening two barter shops. If households are patient enough such that $\beta(1-2\kappa) - \epsilon \geq 0$, then another solution involves opening two monetary shops. In either case all workers will produce (Full Coverage). If $2(1-x)^2(1-\epsilon) \leq \kappa \leq \left(1-2(1-x)^2\right)(1-\epsilon)$, then the unique solution involves opening a single barter shop (Partial Coverage) and if $\left(1-2(1-x)^2\right)(1-\epsilon) > \kappa$, no shop will be opened (No Coverage).

Proof. Case 1: $x < \frac{1}{2}$. We will use a proof by contradiction. Suppose the optimal solution to the planner's problem involves opening at least one barter shop. We will show that by opening only monetary shops ($\lfloor \frac{1}{x} \rfloor$ monetary shops or $\lceil \frac{1}{x} \rceil$ depending on whether κ is larger or smaller than $\left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon)$) the objective function of the planner can be improved. Denote the number of barter and monetary shops in period t respectively by $n^b(t)$ and $n^m(t)$. We will provide an upper bound for the planner's objective function in terms of $n^b(t)$ and $n^m(t)$. Then we will show this upper bound is maximized when $n^b(t) = 0$ and

$n^m(t) \in \left\{\lfloor \frac{1}{x} \rfloor, \lceil \frac{1}{x} \rceil\right\}$. Finally we show that if households are patient enough, the upper bound for $n^b(t) = 0$ and $n^m(t) \in \left\{\lfloor \frac{1}{x} \rfloor, \lceil \frac{1}{x} \rceil\right\}$ can be attained (and hence is a solution to (2.1) subject to (2.2)-(2.7)).

We have introduced an upper bound for the set of households who produce at a barter shop k in period t, which exchanges the pair $\left(\nu^k, \hat{\nu}^k\right)$:

$$hp\left(k, \nu^k; t\right) \cup hp\left(k, \hat{\nu}^k; t\right) \subset S\left(\hat{\nu}^k(t), \nu^k(t)\right) \cup S\left(\nu^k(t), \hat{\nu}^k(t)\right)$$

Therefore at most $2x^2 n^b(t)$ households produce at the $n^b(t)$ barter shops in period t. We have also introduced an upper bound for the set of households who produce at a monetary shop k' in period t, which exchanges the pair $\left(\nu^{k'} = 0, \hat{\nu}^{k'}\right)$:

$$hp\left(k', \hat{\nu}^{k'}; t\right) \subset S\left(0, \hat{\nu}^{k'}(t)\right)$$

Hence at most $xn^m(t)$ households produce at the $n^m(t)$ monetary shops in period t. Moreover, since there is a unit measure of households in the economy, then at most a unit measure of households can produce in a period. The upper bound for the number of households producing in period t is given by:

$$\max\left\{1, xn^m(t) + 2x^2 n^b(t)\right\}$$

Using the manipulated objective function of the planner, (2.8), the upper bound for the objective function is given by:

$$\sum_{t=0}^{\infty} \beta^t \left\{\max\left\{1, xn^m(t) + 2x^2 n^b(t)\right\}(1-\epsilon) - \left(n^b(t) + n^m(t)\right)\kappa\right\} \tag{2.16}$$

If $\kappa \geq x(1-\epsilon)$, then the cost of opening a shop (monetary or barter) exceeds its benefits:

$$\kappa \geq x(1-\epsilon) > 2x^2(1-\epsilon)$$

where the second inequality follows from $x < \frac{1}{2}$. Hence if $\kappa \geq x(1-\epsilon)$, then no shop will be opened. In the rest of the proof we assume $\kappa < x(1-\epsilon)$.

Since $x < \frac{1}{2}$, replacing barter shops with monetary ones will not reduce the objective function. That is

$$\max\left\{1, xn^m(t) + 2x^2n^b(t)\right\} \leq \max\left\{1, x\left(n^m(t) + n^b(t)\right)\right\}.$$

where the equality holds only if $xn^m(t) + 2x^2n^b(t) \geq 1$. If the inequality is strict, then by replacing all the barter shops with monetary shops the planner can increase the objective function, hence contradicting the optimality of the solution. If the equality holds (i.e. $xn^m(t) + 2x^2n^b(t) \geq 1$), and $x\left(n^m(t) + n^b(t) - 1\right) \geq 1$ then the planner can close all the barter shops and increase the number of monetary shops to $n^m(t) + n^b(t) - 1$, that is the total number of shops drops by one. By doing so the value of objective function in period t will increase from $(1-\epsilon) - \left(n^m(t) + n^b(t)\right)\kappa$ to $(1-\epsilon) - \left(n^m(t) + n^b(t) - 1\right)\kappa$. Therefore the only remaining case is

$$xn^m(t) + 2x^2n^b(t) \geq 1 > x\left(n^m(t) + n^b(t) - 1\right) \tag{2.17}$$

which implies $\frac{1}{n^b-1} > x \geq \frac{1}{2} - \frac{1}{2n^b}$, and can only hold if $n^b(t) \leq 3$.

If $n^b(t) = 3$, then (2.17) holds only if $n^m(t) = 0$ and $\frac{1}{3} \leq x < \frac{1}{2}$. If $x \geq \frac{1}{3}$, then 3 monetary shops are enough to cover the taste space and induce all workers to produce provided they are patient enough (we will provide the complete proof below). Although three times the area covered by a barter shop with $\frac{1}{2} > x \geq \frac{1}{3}$ is not necessarily smaller than one (that is $3 \cdot 2x^2 \geq 1$), because of the overlap of the area covered by these three shops, we show they cannot cover the taste space.

Suppose the three barter shops cover the taste space. Denote the pairs of goods traded at these shops by $\left(\nu^1, \hat{\nu}^1\right)$, $\left(\nu^2, \hat{\nu}^2\right)$ and $\left(\nu^3, \hat{\nu}^3\right)$. The shops must cover the diagonal of the taste space $\{(\tau_c, \tau_p) \mid \tau_c = \tau_p \in \ \nu\}$ as well. The diagonal can be divided into three segments such that each segment is covered completely by at least one of the barter shops. Let's denote the length of these segments by $\sqrt{2}a_1$, $\sqrt{2}a_2$ and $\sqrt{2}a_3$. By definition $a_1 + a_2 + a_3 = 1$. Without loss of generality we assume the first, second and third barter shops cover the first, second and third segments respectively. In order to cover $\sqrt{2}a_1$ of the diagonal, the overlap of $S\left(\nu^1, \hat{\nu}^1\right)$ and $S\left(\hat{\nu}^1, \nu^1\right)$ has an area greater than or equal to $(a_1)^2$. Similarly the area of the overlap of pairs of the S sets associated with the second and the third barter shops are greater than or equal to $(a_2)^2$ and $(a_3)^2$. Moreover, since $S\left(\nu^1, \hat{\nu}^1\right)$ and $S\left(\nu^2, \hat{\nu}^2\right)$ cover the first and second segments the area of overlap between them will be at least as large as $(x - a_1)(x - a_2)$. Similarly there will be overlap between the neighboring S sets. The lower bound for the size of overlap areas is given by

$$\begin{aligned}\sum_{i=1}^{3} a_i^2 + \sum_{i=1}^{3}\sum_{j=1} 2\left(x - a_i\right)\left(x - a_j\right) &= \left(\sum_{i=1}^{3} a_i\right)^2 + 6x^2 - 4x\left(\sum_{i=1}^{3} a_i\right) \\ &= 1 + 6x^2 - 4x. \qquad (2.18)\end{aligned}$$

Since the area covered by the six S sets is at most $6x^2$ minus the lower bound of the overlap area given in (2.18), the total area covered by the three barter shops is not greater than $4x - 1 < 1$, where the inequality follows from $x < \frac{1}{2}$. Therefore the three barter shop cannot cover the taste space, while three monetary shops do cover the taste space.

If $n^b(t) = 2$ and (2.17) holds we have

$$(n^m + 2)\, x > 1 > (n^m + 1)\, x$$

where the first inequality follows from $x < \frac{1}{2}$ which implies $2n^b x^2 < 2x$. Therefore, $n^m(t) = \lfloor \frac{1}{x} \rfloor$. As we stated before (2.17) also implies $x > \frac{1}{2} - \frac{1}{2n^b} = \frac{1}{4}$. Hence, $n^m(t)$ can only take the values of 2 (for $\frac{1}{3} < x < \frac{1}{2}$) and 1 (for $\frac{1}{4} < x < \frac{1}{3}$). For $n^m = 2$, the area covered by the monetary shops will be two horizontal strips each of width at most x. Therefore the remaining area should be covered by two barter shops. Since by definition the areas covered by the exchange neighborhoods generated by the two barter shops are symmetric with respect to the diagonal of the type space, the size of the area solely covered by the monetary shops cannot exceed $4x^2$, and the rest of the two strips overlap with the area covered by the barter shops. The area covered by the two barter shops is at most $4x^2$, so in order to cover the type space by the two barter shops and the two monetary shops we should have $8x^2 \geq 1$ which violates $x \leq 3$. For $n^m = 1$, similarly we can show the area covered solely by the monetary shop cannot exceed a square of size x^2 located on the diagonal of the taste space. But then if the type space can be covered by two barter shops and a square of size x^2 on the diagonal, it can also be covered by three barter shops, which contradicts what we showed above.

If $n^b(t) = 1$ and (2.17) holds, then we have $n^m(t) = \lfloor \frac{1}{x} \rfloor$. In this case there is a horizontal strip of the type space which is not covered by the exchange neighborhoods of the monetary shops. The exchange neighborhoods of the barter shop should cover this strip. But since the width of each of them is $x < \frac{1}{2}$, they cannot cover an strip which with the width of 1.

In summary we have shown for $x < \frac{1}{2}$, it is optimal to cover the type space with monetary exchange neighborhoods without opening any barter shop. That is the solution for (2.16) is attained by setting $n^b(t) = 0$. Next we characterize the optimal number of monetary shops, $n^m(t)$ and show the optimal solution satisfies the feasibility and participation constraints of the planner's problem (i.e. (2.2)-(2.7)).

At most x workers can produce at a monetary shop. Hence if $x(1-\epsilon) \leq \kappa$, an operating monetary shop can at most compensate the cost of its operation and the disutility incurred by the workers and does not generate any value added. In this case the optimal solution for the planner's problem is not to open a monetary shop.

If $x(1-\epsilon) > \kappa$, then as long as $(n^m+1)x < 1$ by adding one more monetary shop, the upper bound for the planner's objective, (2.16), increases, and the maximum is attained either by $n^m = \lfloor \frac{1}{x} \rfloor$ or $n^m = \lceil \frac{1}{x} \rceil$. The first one yields the value of $\lfloor \frac{1}{x} \rfloor (x(1-\epsilon) - \kappa)$, and the second one delivers the value of $(1-\epsilon) - \lceil \frac{1}{x} \rceil \kappa$. Hence if $\left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon) < \kappa$ then the benefit of adding one more monetary shop to $\lfloor \frac{1}{x} \rfloor$ of them does not compensate for the increase in the cost, and the upper bound (2.16) is maximized by setting $n^m = \lfloor \frac{1}{x} \rfloor$. Otherwise $n^m = \lceil \frac{1}{x} \rceil$ maximizes the upper bound. Next we show provided households are patient enough, the upper bound can be attained.

Let's assume $\left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon) \geq \kappa$. Hence the upper bound is maximized by $n^m = \lceil \frac{1}{x} \rceil$ with full coverage. Suppose every period the planner opens $\lceil \frac{1}{x} \rceil$ monetary shops exchanging money with goods $\left(i - \frac{1}{2}\right)/\lceil \frac{1}{x} \rceil$ for $i \in \left\{1, 2, \ldots, \lceil \frac{1}{x} \rceil\right\}$, covering the whole taste space. Moreover, in period t the monetary shop trading the pair $\left(0, \left(i - \frac{1}{2}\lceil \frac{1}{x} \rceil\right)\right)$, only exchanges 1 unit of good $\left(i - \frac{1}{2}\lceil \frac{1}{x} \rceil\right)$ for $\overline{m}(t)$ units of money, and $\overline{m}(t)$ units of money for $1 - \lceil \frac{1}{x} \rceil \kappa$ units of good $\left(i - \frac{1}{2}\right)/\lceil \frac{1}{x} \rceil$. The workers and shoppers are directed to the closest shops for production and consumption if they have $\overline{m}(t)$, otherwise they should stay at home. That is, a type $\tau_p \in \left((i-1)/\lceil \frac{1}{x} \rceil, i/\lceil \frac{1}{x} \rceil\right]$ worker is directed to shop i to produce $\ell = 1$ unit of good $\left(i - \frac{1}{2}\right)/\lceil \frac{1}{x} \rceil$, in exchange for $\delta_p = \overline{m}(t)$ units of money, and a type $\tau_c \in \left((j-1)/\lceil \frac{1}{x} \rceil, j/\lceil \frac{1}{x} \rceil\right)$ worker is directed to shop j to exchange $\delta_c = \overline{m}(t)$ units of money to consume $c = 1 - \lceil \frac{1}{x} \rceil \kappa$ units of good $\left(j - \frac{1}{2}\right)\lceil \frac{1}{x} \rceil$. Every period at each shop $1/\lceil \frac{1}{x} \rceil$ units of goods is produced from which κ units is consumed by the shopkeeper, and the remaining $1/\lceil \frac{1}{x} \rceil - \kappa$ units is consumed by the $1/\lceil \frac{1}{x} \rceil$ shoppers who consume at the shop. The shopkeepers do not accumulate money and their participation constraint (2.5) is satisfied since they consume κ every period they are open. On the equilibrium path households start and end period t with $\overline{m}(t)$ units of money. Finally by producing one unit in period t, a household incurs disutility $-\epsilon$ and receives money which can be used to consume $1 - \lceil \frac{1}{x} \rceil \kappa$ units of its desired good next period.

Since $\left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon) \geq \kappa$ we have $\lceil \frac{1}{x} \rceil = \lfloor \frac{1}{x} \rfloor + 1$, hence $1/\left(1 - \lfloor \frac{1}{x} \rfloor x\right) > \lceil \frac{1}{x} \rceil$. Therefore

$$(1-\epsilon) > \frac{\kappa}{1 - \lfloor \frac{1}{x} \rfloor x} > \lceil \frac{1}{x} \rceil \kappa$$

which implies

$$c = 1 - \lceil \frac{1}{x} \rceil \kappa > \epsilon.$$

Now if β is sufficiently high we also have $c = 1 - \lceil \frac{1}{x} \rceil \kappa > \frac{\epsilon}{\beta}$, and household participation constraint (2.6) holds.

If $x(1-\epsilon) > k > \left(1 - \lfloor \frac{1}{x} \rfloor x\right)(1-\epsilon)$, hence the upper bound is maximized by $n^m = \lfloor \frac{1}{x} \rfloor$ with partial coverage. Suppose, starting from period $t = 0$ every even period the planner opens $\lfloor \frac{1}{x} \rfloor$ monetary shops exchanging money with goods $\frac{i-1}{\lfloor \frac{1}{x} \rfloor} + \frac{x}{2}$ for $i \in \left\{1, 2, \ldots, \lfloor \frac{1}{x} \rfloor\right\}$, and every odd period the planner opens $\lfloor \frac{1}{x} \rfloor$ monetary shops exchanging money with goods $\frac{i}{\lfloor \frac{1}{x} \rfloor} - \frac{x}{2}$ for $i \in \left\{1, 2, \ldots, \lfloor \frac{1}{x} \rfloor\right\}$.

Households are endowed with $\eta(0) = \overline{m}(0)$ units of money. The supply of money grows at rate γ by a type independent lump sum money transfer of $\eta(t) = \gamma(1-\gamma)^{t-1}\overline{m}(0)$ for $t \geq 1$. Every period t there will be $\lfloor \frac{1}{x} \rfloor x$ of workers who produce and receive $\bar{\delta}_p(t)$ units of money. Planner's proposals to the households in the even periods are summarized as below, where $\tau_c \in \left(\frac{i-1}{\lfloor \frac{1}{x} \rfloor}, \frac{i}{\lfloor \frac{1}{x} \rfloor}\right]$ shoppers are directed to shop i when $c > 0$ and $\tau_p \in \left(\frac{j-1}{\lfloor \frac{1}{x} \rfloor}, \frac{j}{\lfloor \frac{1}{x} \rfloor}\right]$ workers are directed to shop j when $\ell = 1$.

Table A.1

$\tau_p \backslash \tau_c$	$\left(\frac{i-1}{\lfloor\frac{1}{x}\rfloor}, \frac{i}{\lfloor\frac{1}{x}\rfloor} - x\right)$	$\left[\frac{i}{\lfloor\frac{1}{x}\rfloor} - x, \frac{i-1}{\lfloor\frac{1}{x}\rfloor} + x\right]$	$\left(\frac{i-1}{\lfloor\frac{1}{x}\rfloor} + x, \frac{i}{\lfloor\frac{1}{x}\rfloor}\right]$
$\left(\frac{j-1}{\lfloor\frac{1}{x}\rfloor} + x, \frac{j}{\lfloor\frac{1}{x}\rfloor}\right]$	$\ell = 0$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = 0$ $\delta_c = \bar{\delta}_p(-1)$ $+\eta + \eta(-1)$	$\ell = 0$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = 0$ $\delta_c = \bar{\delta}_p(-1)$ $+\eta + \eta(-1)$	$\ell = 0$ $c = 0$ $\delta_p = 0$ $\delta_c = 0$
$\left[\frac{j}{\lfloor\frac{1}{x}\rfloor} - x, \frac{j-1}{\lfloor\frac{1}{x}\rfloor} + x\right]$	$\ell = 1$ $c = 2\left(1 - \frac{\kappa}{x}\right)$ $\delta_p = \bar{\delta}_p$ $\delta_c = \bar{\delta}_p(-1) + \bar{\delta}_p(-2)$ $+\eta + \eta(-1)$	$\ell = 1$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = \bar{\delta}_p$ $\delta_c = \bar{\delta}_p(-1) + \eta$	$\ell = 1$ $c = 0$ $\delta_p = \bar{\delta}_p$ $\delta_c = 0$
$\left(\frac{j-1}{\lfloor\frac{1}{x}\rfloor}, \frac{j}{\lfloor\frac{1}{x}\rfloor} - x\right)$	$\ell = 1$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = \bar{\delta}_p$ $\delta_c = \bar{\delta}_p(-2)$ $+\eta + \eta(-1)$	$\ell = 1$ $c = 0$ $\delta_p = \bar{\delta}_p$ $\delta_c = 0$	$\ell = 1$ $c = 0$ $\delta_p = \bar{\delta}_p$ $\delta_c = 0$

where $z(-1)$ denotes $z(t-1)$ (e.g. $\eta(-1)$ denotes $\eta(t-1)$).

For the odd period, planner proposals are summarized as follows.

Table A.2

$\tau_p \backslash \tau_c$	$\left(\frac{i-1}{\lfloor\frac{1}{x}\rfloor}, \frac{i}{\lfloor\frac{1}{x}\rfloor} - x\right)$	$\left[\frac{i}{\lfloor\frac{1}{x}\rfloor} - x, \frac{i-1}{\lfloor\frac{1}{x}\rfloor} + x\right]$	$\left(\frac{i-1}{\lfloor\frac{1}{x}\rfloor} + x, \frac{i}{\lfloor\frac{1}{x}\rfloor}\right]$
$\left(\frac{j-1}{\lfloor\frac{1}{x}\rfloor} + x, \frac{j}{\lfloor\frac{1}{x}\rfloor}\right]$	$\ell = 1$ $c = 0$ $\delta_p = \bar{\delta}_p$ $\delta_c = 0$	$\ell = 1$ $c = 0$ $\delta_p = \bar{\delta}_p$ $\delta_c = 0$	$\ell = 1$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = \bar{\delta}_p$ $\delta_c = \bar{\delta}_p(-2)$ $+\eta + \eta(-1)$
$\left[\frac{j}{\lfloor\frac{1}{x}\rfloor} - x, \frac{j-1}{\lfloor\frac{1}{x}\rfloor} + x\right]$	$\ell = 1$ $c = 0$ $\delta_p = \bar{\delta}_p$ $\delta_c = 0$	$\ell = 1$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = \bar{\delta}_p$ $\delta_c = \bar{\delta}_p(-1) + \eta$	$\ell = 1$ $c = 2\left(1 - \frac{\kappa}{x}\right)$ $\delta_p = \bar{\delta}_p$ $\delta_c = \bar{\delta}_p(-1) + \bar{\delta}_p(-2)$ $+\eta + \eta(-1)$
$\left(\frac{j-1}{\lfloor\frac{1}{x}\rfloor}, \frac{j}{\lfloor\frac{1}{x}\rfloor} - x\right)$	$\ell = 0$ $c = 0$ $\delta_p = 0$ $\delta_c = 0$	$\ell = 0$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = 0$ $\delta_c = \bar{\delta}_p(-1)$ $+\eta + \eta(-1)$	$\ell = 0$ $c = 1 - \frac{\kappa}{x}$ $\delta_p = 0$ $\delta_c = \bar{\delta}_p(-1)$ $+\eta + \eta(-1)$

where the transfer of money upon production, $\bar{\delta}_p$ evolves according to:

$$\begin{aligned} \bar{\delta}_p(t) &= \lfloor\frac{1}{x}\rfloor x \bar{\delta}_p(t-1) + \left(1 - \lfloor\frac{1}{x}\rfloor x\right) \bar{\delta}_p(t-2) \\ &\quad + \left(\frac{1}{\lfloor\frac{1}{x}\rfloor x} - 2\left(1 - \lfloor\frac{1}{x}\rfloor x\right)\right) \eta(t) + 2\left(1 - \lfloor\frac{1}{x}\rfloor x\right) \eta(t-1). \end{aligned} \tag{2.19}$$

Therefore, the shopkeepers do not accumulate money along the equilibrium path.

Household participation constraints (2.6) can be summarized for each region of the taste space as follows:

$\tau_p \backslash \tau_c$	$\left(\frac{i-1}{\lfloor\frac{1}{x}\rfloor}, \frac{i}{\lfloor\frac{1}{x}\rfloor} - x\right)$	$\left[\frac{i}{\lfloor\frac{1}{x}\rfloor} - x, \frac{i-1}{\lfloor\frac{1}{x}\rfloor} + x\right]$	$\left(\frac{i-1}{\lfloor\frac{1}{x}\rfloor} + x, \frac{i}{\lfloor\frac{1}{x}\rfloor}\right]$
$\left(\frac{j-1}{\lfloor\frac{1}{x}\rfloor} + x, \frac{j}{\lfloor\frac{1}{x}\rfloor}\right]$	$\beta\left(1-\frac{\kappa}{x}\right) \geq \epsilon$	$\beta\left(1-\frac{\kappa}{x}\right) \geq \epsilon$	$\beta^2\left(1-\frac{\kappa}{x}\right) \geq \epsilon$
$\left[\frac{j}{\lfloor\frac{1}{x}\rfloor} - x, \frac{j-1}{\lfloor\frac{1}{x}\rfloor} + x\right]$	$\beta^2 2\left(1-\frac{\kappa}{x}\right) \geq (1+\beta)\epsilon$	$\beta\left(1-\frac{\kappa}{x}\right) \geq \epsilon$	$\beta^2 2\left(1-\frac{\kappa}{x}\right) \geq (1+\beta)\epsilon$
$\left(\frac{j-1}{\lfloor\frac{1}{x}\rfloor}, \frac{j}{\lfloor\frac{1}{x}\rfloor} - x\right)$	$\beta^2\left(1-\frac{\kappa}{x}\right) \geq \epsilon$	$\beta\left(1-\frac{\kappa}{x}\right) \geq \epsilon$	$\beta\left(1-\frac{\kappa}{x}\right) \geq \epsilon$

Since $x(1-\epsilon) > \kappa$, we have $1 - \frac{\kappa}{x} > \epsilon$, and if β is high enough (i.e. such that $\beta^2\left(1-\frac{\kappa}{x}\right) \geq \epsilon$), then the household participation constraint will hold for all regions of the taste space. In summary, with partial coverage, x units of workers will produce at each monetary shop every period. From this x units of goods produced at each monetary shop, κ units of them will be consumed by the shopkeeper satisfying the shopkeeper participation constraint (2.5). The remaining $x - \kappa$ units of goods at each shop will be consumed by the $\frac{1}{\lfloor\frac{1}{x}\rfloor} - 2x\left(1 - \lfloor\frac{1}{x}\rfloor x\right)$ shoppers visiting the shop, from which $x - 2\left(2x\lfloor\frac{1}{x}\rfloor - 1\right)\left(\frac{1}{\lfloor\frac{1}{x}\rfloor} - x\right)$ of those who are consuming $\left(1 - \frac{\kappa}{x}\right)$ units after having produced only since their last consumption and $\left(2x\lfloor\frac{1}{x}\rfloor - 1\right)\left(\frac{1}{\lfloor\frac{1}{x}\rfloor} - x\right)$ of those who are consuming $2\left(1-\frac{\kappa}{x}\right)$ units after having produced twice since their last consumption. Figure (2.4) illustrates the location of monetary shops over time and the set of shoppers and workers who consume and produce when $\lfloor\frac{1}{x}\rfloor = 2$.

Case 2: $\frac{1}{2} \leq x < \frac{2}{3}$. With a single monetary shop at most x units of workers can produce. Two monetary shops provide production opportunities for all of the workers, fully covering the type space.

A single barter shop can cover $1-2(1-x)^2$ units of the type space. For example if a barter shop is trading $\left(\frac{x}{2}, 1-\frac{x}{2}\right)$ then the households with type $(\tau_c, \tau_p) \in [0,x] \times [1-x, 1]$ can produce good $1-\frac{x}{2}$ in exchange for good $\frac{x}{2}$, and the households with type $(\tau_c, \tau_p) \in [1-x, 1] \times [0, x]$ can produce good $\frac{x}{2}$ in exchange for good $1-\frac{x}{2}$, provided the participation constraint holds. Notice that a single barter shop has two exchange neighborhoods each covering x^2 units of the type space. But for $x \geq \frac{1}{2}$, the two neighborhoods will have an overlap with the minimum size of $(2x-1)^2$. Hence the area of type space covered by a

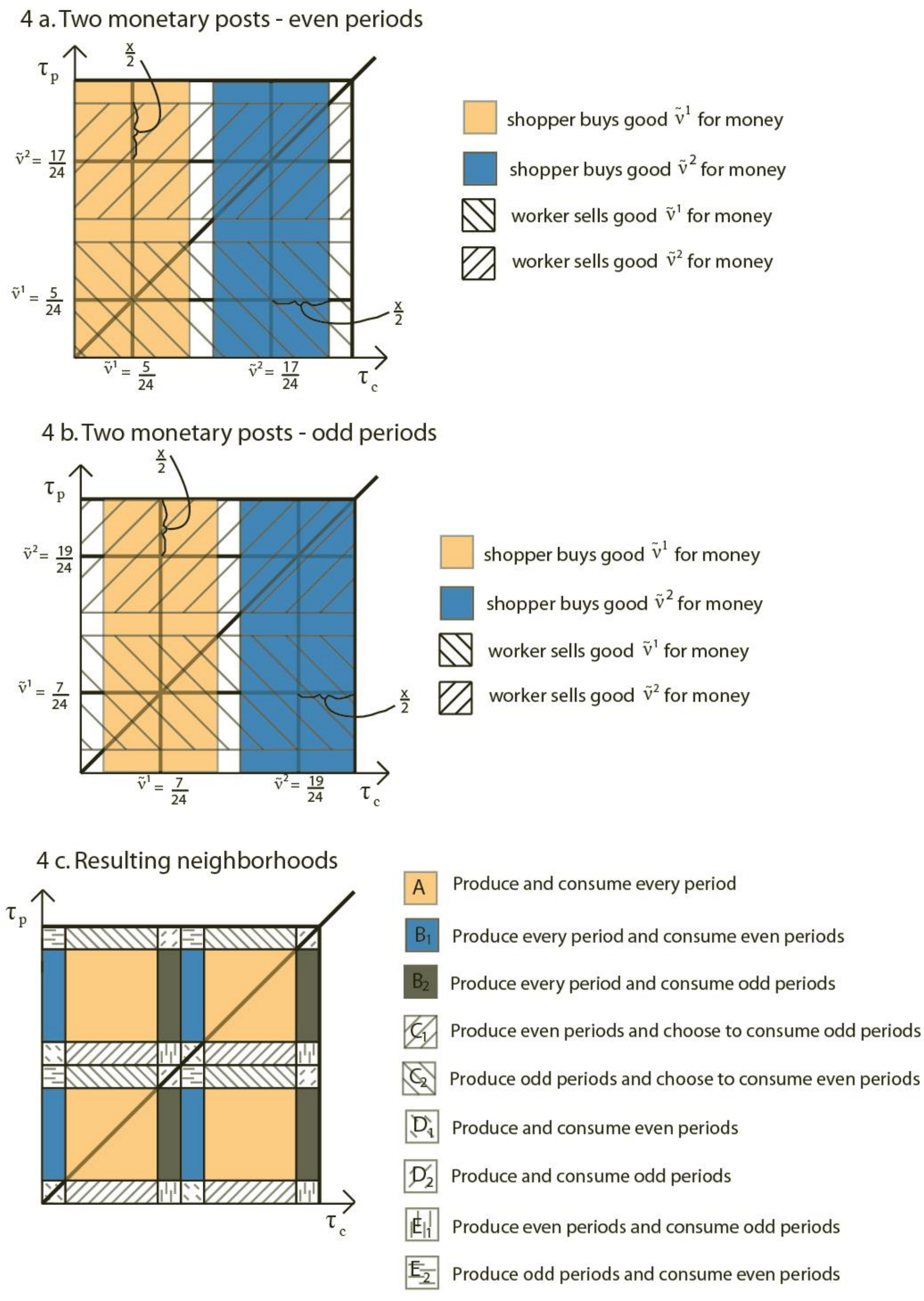


Figure 2.4: Moving Monetary Exchange Neighborhoods

single barter shop will have the maximum size of

$$2x^2 - (2x-1)^2 = 1 - 2(1-x)^2.$$

We show for $x < \frac{2}{3}$ two barter shops cannot cover the type space using a proof by contradiction. Suppose the type space is covered by the exchange neighborhoods of the two barter shops. Since $x < \frac{2}{3}$ one barter shop cannot cover more than $\frac{2}{3}$ of the diagonal of the type space. Hence each barter shop should at least cover $\frac{1}{3}$ of the diagonal. Without loss of generality we can assume the first barter shop is trading the pair $\left(\nu^1, \hat{\nu}^1 = 1 - \nu^1\right)$ of goods[29]. The exchange neighborhood of this shop has to cover the middle $\frac{1}{3}$ of the type space's diagonal; that is the households with type $\left(\frac{1}{3}, \frac{1}{3}\right)$ and $\left(\frac{2}{3}, \frac{2}{3}\right)$ should be covered by the first shop. Therefore $|\nu^1 - \frac{1}{3}| \leq \frac{x}{2} < \frac{1}{3}$ and $|\frac{2}{3} - \nu^1| \leq \frac{x}{2} < \frac{1}{3}$, which also imply $\frac{1}{3} \leq \nu^1 \leq \frac{2}{3}$. This leaves the households with type $\left(\frac{1}{6}, \frac{1}{6}\right)$, $\left(\frac{5}{6}, \frac{5}{6}\right)$, $\left(\frac{1}{2}, 1\right)$ and $\left(\frac{1}{2}, 1\right)$ uncovered by the first shop's exchange neighborhoods, so they should be covered by the second barter shop. If the second shop is covering $\left(\frac{1}{6}, \frac{1}{6}\right)$ and $\left(\frac{5}{6}, \frac{5}{6}\right)$ types, then

$$|\frac{1}{6} - \nu^2|_{mod1} \leq \frac{x}{2} < \frac{1}{3}$$

imply

$$\left\{\left(\nu^2, \hat{\nu}^2\right), \left(\hat{\nu}^2, \nu^2\right)\right\} \in (0, \frac{1}{6}]^2 \cup [\frac{5}{6}, 1]^2 \cup (0, \frac{1}{6}] \times [\frac{5}{6}, 1] \cup [\frac{5}{6}, 1] \times (0, \frac{1}{6}]$$

but then the households with type $\left(\frac{1}{2}, 1\right)$ and $\left(\frac{1}{2}, 1\right)$ cannot be directed to the second barter shop for production and consumption since $x < \frac{2}{3}$, thereby yielding the contradiction (See Figure (2.3)c for the areas on the border which are not covered).

The upper bound for the planner's objective function (2.8) given by (2.10), is equal to $\frac{1}{(1-\beta)}\left\{\left(1 - 2(1-x)^2\right)(1-\epsilon) - \kappa\right\}$ with a single barter shop, is equal to $\frac{1}{(1-\beta)}\left\{x(1-\epsilon) - \kappa\right\}$ with one monetary shop, and is equal to $\frac{1}{(1-\beta)}\left\{(1-\epsilon) - 2\kappa\right\}$ with two monetary shops.

[29] This can be shown by transforming the coordinates of the type space along the diagonal.

Hence with $\kappa > \left(1-2(1-x)^2\right)(1-\epsilon)$, the planner does not open any shop. If $\left(1-2(1-x)^2\right)(1-\epsilon) > \kappa > 2(1-x)^2(1-\epsilon)$, then the upper bound is given by a single barter shop and for $2(1-x)^2(1-\epsilon) > \kappa$ two monetary shops yield the maximum of the upper bound.

If $\left(1-2(1-x)^2\right)(1-\epsilon) > \kappa > 2(1-x)^2(1-\epsilon)$, with a single barter shop trading the pair $\left(\frac{x}{2}, 1-\frac{x}{2}\right)$ of goods, $1-2(1-x)^2$ workers produce $1-2(1-x)^2$ units of good from which κ unit will be consumed by the shopkeeper, satisfying the shopkeeper's participation constraint (2.5), and the remaining $1-2(1-x)^2-\kappa$ units will be consumed by the $1-2(1-x)^2$ shoppers visiting the shop. If the utility from consumption compensates the production disutility, that is

$$1-\frac{\kappa}{1-2(1-x)^2} \geq \epsilon$$

or equivalently $\left(1-2(1-x)^2\right)(1-\epsilon) \geq \kappa$, then the household participation constraint (2.6) is satisfied.

If $\kappa < 2(1-x)^2(1-\epsilon)$ with two monetary shops exchanging money with goods $\frac{1}{4}$ and $\frac{3}{4}$, all households can produce. In period t workers of type $\tau_p \in (\frac{i-1}{2}, \frac{i}{2}]$ are directed to the monetary shop $i \in \{1,2\}$ to trade 1 unit of good $\frac{i-1}{2}+\frac{1}{4}$ with $\overline{m}(t)$ units of money. The shoppers of type $\tau_c \in (\frac{i-1}{2}, \frac{i}{2}]$ are directed to the monetary shop $i \in \{1,2\}$ to trade $\overline{m}(t)$ units of money with $1-2\kappa$ unit of good $\frac{i-1}{2}+\frac{1}{4}$. From the $\frac{1}{2}$ units of good produced at each shop, κ units are consumed by the shopkeepers, satisfying the participation constraint of the shopkeeper (2.5), and the remaining will be consumed by the visiting shoppers. Feasibility constraints (2.2)-(2.3) hold and shopkeepers do not accumulate money. Since $x \geq \frac{1}{2}$, the condition $\kappa < 2(1-x)^2(1-\epsilon)$ implies $1-2\kappa > \epsilon$. Finally, if households are patient enough (i.e. such that $1-2\kappa > \frac{\epsilon}{\beta}$), the household participation constraint (2.6) holds.

Case 3: $\frac{2}{3} \leq x$. This case is similar to the previous one, except that two barter shops exchanging the pairs $\left(\frac{1}{2}, \frac{1}{2}\right)$ and $\left(\frac{1}{6}, \frac{5}{6}\right)$ can also cover the whole type space. Hence if $\kappa < 2(1-x)^2(1-\epsilon)$ and households are sufficiently patient, the solution is not unique; the planner can choose to open two barter shops or two monetary shops to cover the type space. If households are not patient enough, then the unique solution involves opening two barter shops. ■

Lemma 1. For $x \in \left(\frac{1}{3}, \frac{1}{2}\right)$, the size of the overlap area among the exchange neighborhoods of two barter shops is greater than or equal to $(3x-1)^2$.

Proof. Denote the two barter shops 1 and 2 trading the pairs $\left(\nu^1, \hat{\nu}^1\right)$ and $\left(\nu^2, \hat{\nu}^2\right)$, where as before we have $\nu^1 \leq \hat{\nu}^1$, $\nu^2 \leq \hat{\nu}^2$ and we also assume $\nu^1 \leq \nu^2$ and $|\hat{\nu}^2 - \nu^2|_{mod1} \leq |\hat{\nu}^1 - \nu^1|_{mod1}$, without loss of generality. Changing the coordinates of the type space using $(mod1)$ arithmetic, without loss of generality, we can assume $\hat{\nu}^2 = 1 - \nu^2$ and $\nu^2 \geq \frac{1}{4}$, so that $|\hat{\nu}^2 - \nu^2|_{mod1} = 1 - 2\nu^2$.[30] Notice that despite the coordinates change and $|\hat{\nu}^2 - \nu^2|_{mod1} \leq |\hat{\nu}^1 - \nu^1|_{mod1}$, we can still assume $\nu^1 \leq \nu^2$; if $\nu^1 > \nu^2$, then replacing the $\left(\nu^1, \hat{\nu}^1\right)$ with $\left(1 - \hat{\nu}^1, 1 - \nu^1\right)$ will not affect the size of the overlap area.

For the two barter shops the exchange neighborhoods generate 4 squares. Denote the squares of the exchange neighborhoods around $\left(\nu^1, \hat{\nu}^1\right)$, $\left(\hat{\nu}^1, \nu^1\right)$, $\left(\nu^2, \hat{\nu}^2\right)$ and $\left(\hat{\nu}^2, \nu^2\right)$ respectively by S_1, $S_{1'}$, S_2 and $S_{2'}$. These squares can have 6 areas of overlap among them. The size of the overlap area is given by:

$$\begin{aligned}
&\max(0, x - |\hat{\nu}^1 - \nu^1|_{mod1})^2 + \max(0, x - |\hat{\nu}^2 - \nu^2|_{mod1})^2 \\
&+2\max(0, x - |\nu^2 - \nu^1|_{mod1})\max(0, x - |\hat{\nu}^2 - \hat{\nu}^1|_{mod1}) \\
&+2\max(0, x - |\hat{\nu}^2 - \nu^1|_{mod1})\max(0, x - |\hat{\nu}^1 - \nu^2|_{mod1}). \qquad (2.20)
\end{aligned}$$

If $x - |\hat{\nu}^1 - \nu^1|_{mod1} \geq (3x-1)$ or $x - |\hat{\nu}^2 - \nu^2|_{mod1} \geq (3x-1)$, we are done. Hence in the remaining cases we have

$$\begin{aligned}
1 - 2x < \hat{\nu}^1 - \nu^1 &< 2x \qquad (2.21) \\
\frac{1}{2} - x < \nu^2 &< x \qquad (2.22)
\end{aligned}$$

[30] This can be done by adding or subtracting $\frac{1}{2}\left(1 - \hat{\nu}^2 - \nu^2\right)$ to all coordinates in $(mod1)$.

where the second set of inequalities follows from $\hat{\nu}^2 = 1-\nu^2$.

If S_1 does not overlap with $S_{2'}$, then $\nu^1+\nu^2 \geq x$. [31] Hence $\left(x-\hat{\nu}^2+\hat{\nu}^1\right) = \left(x-(1-\nu^2)+\hat{\nu}^1\right) \geq 2x-1+\hat{\nu}^1-\nu^1 > 0$, where the third inequality follows from (2.21). Moreover, (2.22) implies $\left(x-\nu^2+\nu^1\right) > 0$, thereby implying S_1 and S_2 overlap. The size of the overlap area (2.20) in this case is given by:

$$\begin{aligned}
&\max(0, x-|\hat{\nu}^1-\nu^1|_{mod1})^2+\max(0, x-|\hat{\nu}^2-\nu^2|_{mod1})^2 \\
&+2\left(x-1+\nu^2+\hat{\nu}^1\right)\left(x-\nu^2+\nu^1\right)
\end{aligned}$$

for $\hat{\nu}^2 = 1-\nu^2$. From $\nu^1+\nu^2 \geq x$ it follows that

$$\begin{aligned}
(3x-1) &\leq (x-(\hat{\nu}^1-\nu^1))+\left(x-(1-\nu^2)+\hat{\nu}^1\right) \\
&\leq \max(0, x-|\hat{\nu}^1-\nu^1|_{mod1})+\left(x-(1-\nu^2)+\hat{\nu}^1\right)
\end{aligned}$$

and

$$\begin{aligned}
(3x-1) &\leq (x-((1-\nu^2)-\nu^2))+\left(x-\nu^2+\nu^1\right) \\
&\leq \max(0, x-|(1-\nu^2)-\nu^2|_{mod1})+\left(x-\nu^2+\nu^1\right).
\end{aligned}$$

Therefore

$$\max(0, x-|\hat{\nu}^1-\nu^1|_{mod1})^2+\max(0, x-|\hat{\nu}^2-\nu^2|_{mod1})^2+2\left(x-1+\nu^2+\hat{\nu}^1\right)\left(x-\nu^2+\nu^1\right)$$

$$\begin{aligned}
\geq\ &\max(0, x-|\hat{\nu}^1-\nu^1|_{mod1})^2+\max(0, x-|\hat{\nu}^2-\nu^2|_{mod1})^2 \\
&+2((3x-1)-\max(0, x-|\hat{\nu}^1-\nu^1|_{mod1}))
\end{aligned}$$

[31] If S_1 does not overlap with $S_{2'}$, alternatively, we might have $\nu^1+\nu^2 < x$ and instead $\hat{\nu}^1 > \nu^2+x$ would guarantee the separation of S_1 and $S_{2'}$. But these two conditions together imply $\hat{\nu}^2-\nu^2 = 1-2\nu^2 > 1-\hat{\nu}^1+\nu^1$, which contradicts $|\hat{\nu}^2-\nu^2|_{mod1} \leq |\hat{\nu}^1-\nu^1|_{mod1}$.

$$\times((3x-1)-\max(0,x-|\hat{\nu}^2-\nu^2|_{mod1}))\geq(3x-1)^2.$$

If S_1 does not overlap with S_2, since $\nu^2-\nu^1<x$ from (2.22), we have $\hat{\nu}^2-\hat{\nu}^1=1-\nu^2-\hat{\nu}^1\geq x$. [32] Hence we have $1-\nu^2-\hat{\nu}^1>\nu^2-\nu^1$and therefore $1-2\nu^2>\hat{\nu}^1-\nu^1$, which contradicts $|\hat{\nu}^2-\nu^2|_{mod1}\leq|\hat{\nu}^1-\nu^1|_{mod1}$.

The only remaining case is when S_1 overlaps with both S_2 and $S_{2'}$. In this case the size of the overlap area (2.20) is given by:

$$\begin{aligned}
&\max(0,x-|\hat{\nu}^1-\nu^1|_{mod1})^2+\max(0,x-|\hat{\nu}^2-\nu^2|_{mod1})^2\\
&+2\left(x-\hat{\nu}^2+\hat{\nu}^1\right)\left(x-\nu^2+\nu^1\right)\\
&+2\left(x-(1+\nu^1)+\hat{\nu}^2\right)\left(x-\hat{\nu}^1+\nu^2\right)
\end{aligned}\tag{2.23}$$

Since $x-\hat{\nu}^1+\nu^2>x-\nu^2+\nu^1$ follows from $1-2\nu^2>1-\hat{\nu}^1+\nu^1\geq1-\hat{\nu}^1-\nu^1$, as long as $\hat{\nu}^2-\nu^2>x$ the overlap area (2.23) can be reduced by decreasing $\hat{\nu}^2$. Therefore in order to prove the size of the overlap area is at least $(3x-1)^2$, it is sufficient to show it when $\hat{\nu}^2-\nu^2\leq x$ or equivalently $1-2\nu^2\leq x$. Similarly it is sufficient to show (2.23) is greater than or equal to $(3x-1)^2$, for $\hat{\nu}^1-\nu^1\leq x$. Given $1-2\nu^2\leq x$ and $\hat{\nu}^1-\nu^1\leq x$, the size of the overlap area (2.23) can be written as:

$$\begin{aligned}
&(x-(\hat{\nu}^1-\nu^1))^2+(x-(1-2\nu^2))^2\\
&+2\left(x-1+\nu^2+\hat{\nu}^1\right)\left(x-\nu^2+\nu^1\right)\\
&+2\left(x-\nu^1-\nu^2\right)\left(x-\hat{\nu}^1+\nu^2\right).
\end{aligned}\tag{2.24}$$

[32] Notice that since $\nu^2<x$, we cannot have $\hat{\nu}^1-\hat{\nu}^2\geq x$.

Now the following four equalities and two inequalities

$$\begin{aligned}
(x-(\hat{\nu}^1-\nu^1))+\left(x-1+\nu^2+\hat{\nu}^1\right)+\left(x-\nu^1-\nu^2\right) &= (3x-1), && (2.25)\\
(x-(1-2\nu^2))+\left(x-\nu^2+\nu^1\right)+\left(x-\nu^1-\nu^2\right) &= (3x-1), && (2.26)\\
\left(x-\hat{\nu}^1+\nu^2\right)+\left(x-\nu^2+\nu^1\right)-(x-(\hat{\nu}^1-\nu^1)) &= x>(3x-1), && (2.27)\\
\left(x-\hat{\nu}^1+\nu^2\right)+\left(x-1+\nu^2+\hat{\nu}^1\right)-(x-(1-2\nu^2)) &= x>(3x-1), && (2.28)
\end{aligned}$$

guarantee that (2.24) is greater than or equal to $(3x-1)^2$. That is by substituting $(x-(\hat{\nu}^1-\nu^1))$ from (2.25), and $\left(x-\nu^2+\nu^1\right)$ from (2.26) into (2.24) and rearranging it, the size of the overlap area can be written as

$$\begin{aligned}
&\left[(3x-1)-\left(x-\nu^1-\nu^2\right)\right]^2+\\
&\left[(x-1+\nu^2+\hat{\nu}^1)-(x-(1-2\nu^2))\right]^2+\\
&2(x-\nu^1-\nu^2)(x-\hat{\nu}^1+\nu^2)
\end{aligned}$$

which in turn, following (2.28), is greater than

$$\left[(3x-1)-\left(x-\nu^1-\nu^2\right)\right]^2+\left[(3x-1)-\left(x-\hat{\nu}^1+\nu^2\right)\right]^2$$

$$+2\left(x-\nu^1-\nu^2\right)\left(x-\hat{\nu}^1+\nu^2\right)\geq(3x-1)^2.$$

■

Chapter 3

A Dynamic Duopoly with Endogenous Horizontal and Vertical Differentiation

3.1 Introduction

This chapter is a study of a dynamic duopoly in which firms may be differentiated in two dimensions: vertical and horizontal. This differentiation is endogenous, in the sense that firms may invest in improving their quality and in adjusting their location over time. Differences between firms' qualities determine vertical differentiation, differences between locations determined horizontal differentiation. We set up an infinite-horizon dynamic model which allows firms to choose investment in both dimensions. After solving for the Markov-perfect equilibrium investment choices we study the effects of various parameter changes on equilibrium outcomes.

Examples of this type of competition may be found in various industries. Manufacturers of personal computers, for example, may differentiate their products from the competition in two ways. First, they invest in improving quality, represented by speed, memory, etc. Second, on occasions a manufacturer may change the architecture of a computer's processor, either by moving it closer to the architecture of its rival, in order to steal the rival's business, or by moving it further away in order to develop local market power. These differentiation choices are of course interdependent – shifts in relative qualities induce changes in architecture, and *vice versa*. The aim of this paper is to examine the fundamental factors underlying these dynamic linkages between quality and architecture.[1]

In the computer industry major shifts in processor architecture are relatively infrequent. Similarly in industries such as retailing it could be expensive for a chain to change the locations of all its stores. However in other industries and other dimensions firms' repositioning of their products could be more frequent. For example a retail clothing chain can adjust its location in the space of consumer tastes by re-stocking with different fashions. At the same time it can invest in quality upgrades, not just in the clothing sold, but also in terms of staffing, store amenities, training, and so on. Or consider radio stations choosing formats for their broadcasting content. The different options – talk, country music, hip-hop, etc. – represent horizontal differentiation; switches in these formats are not uncommon in radio markets. Simultaneously a station may consider upgrading aspects of its overall quality such as local news and traffic, websites, broadcast technology, etc. In turn the station's (or clothing store's) choices in these dimensions will induce responses from its rivals. Over time the pattern of changes in product positioning and quality among competing firms will play out in complicated ways that may not be obvious at first glance.

A particular question of interest here concerns the relationship between competition and the amount of innovation in an industry. Research on the determinants and consequences of innovation has a long heritage in economics, extending back at least to the work of Schumpeter, and encompassing for example the study of patent races (e.g., Reinganum

[1]In 2005 Apple Computer announced the migration of its Macintosh product line from PowerPC processors to the Intel x86 architecture that is the standard platform for Microsoft Windows. With this change Macintosh users will for the first time be able to boot Windows directly on their machines (rather than in slower emulation mode), and hence will have much better access to Windows-compatible software. (http://en.wikipedia.org/wiki/Apple_Intel_Transition , accessed 4 April 2006.) See Bresnahan and Greenstein (1999) for more on the relationship between quality and horizontal differentiation in the computer industry.

(1982)) and innovation and growth (Grossman and Helpman (1991), Aghion and Howitt (1992)).[2] In recent work Aghion, Harris, Howitt and Vickers (2001) and Aghion, Bloom, Blundell, Griffith and Howitt (2005) have focused on the role of innovation in oligopoly, taking up a question previously addressed in the literature on patent races, and in, e.g., Budd, Harris and Vickers (1993).[3] Aghion et al. (2001), for example, embed a dynamic 'quality-ladder' model of innovation into a duopoly in which the competitiveness of the industry is parameterized by the elasticity of substitution between the two firms' products.[4] Their results suggest an 'inverse-U shaped' relationship between competition and innovation. Starting from a situation of dual monopolies (i.e., a high degree of market power for each firm), as the elasticity of substitution increases the average amount of steady-state innovation (i.e., investment in quality) initially rises, as firms respond to the new competition by attempting to get ahead in the quality race. Eventually however this increase in intensity of investment creates a countervailing effect: the industry spends more and more time in states of asymmetric qualities (where aggregate innovative activity is relatively low) than in states of neck-and-neck competition (where innovative activity is intense). As a result if the degree of competition increases far enough then the long-run average innovation rate may start to fall.

This picture of an 'inverse-U' relationship is borne out by a variant of this duopolistic quality-ladder model employed in Aghion et al. (2005), and supported there by empirical evidence on patent activity in the United Kingdom. In the present paper we take the view that oligopolistic competition has horizontal as well as vertical dimensions and that it is worth studying innovation in a context which makes firms' locations in both these dimensions explicit. That is, the degree of competition in an industry is itself an endogenous variable, reflecting particular features of the environment such as costs of entry, costs of differentiation, and the heterogeneity in consumer tastes. It would be desirable to study

[2]See also Scotchmer (2004), Jaffe and Lerner (2004), etc.

[3]See also Doraszelski (2003).

[4]Here we interpret their model as one of quality improvement, although in their framework firms actually invest in cost reduction. There should be a one-to-one mapping of results from one interpretation to the other.

innovation in a model of oligopoly which allowed for all these features and for a variable number of firms with dynamic entry and exit. For tractability we deal here with the simpler duopoly case, leaving a more general model for future work.

To address the dynamic links between innovation and the two dimensions of differentiation we set up a discrete-time, infinite-horizon dynamic Hotelling model in which firms may invest in each period both in changing architecture (location) and in improving quality. Investment outcomes are stochastic, there is no entry or exit, and for simplicity firms may only locate at the endpoints of the Hotelling line. Competition in quality upgrades takes the form of a quality ladder in which firms' innovative successes produce limited spillovers for their rivals. These spillovers ensure that the 'quality lag' suffered by any such rival remains bounded over time – this gives any laggard a reason to stay in the market. We seek symmetric Markov-perfect equilibria (MPE's) in pure strategies for the game in which firms selling a perishable good compete in prices in each period, taking as given their current configuration of qualities and locations. The general model is too complicated to solve for the MPE's analytically. Hence we start by solving a simpler version of the model with deterministic investment outcomes and fixed quality differences. Solving for the MPE of this simpler model is a non-trivial task, one that is of interest in its own right as a natural extension of the Hotelling model. As far as we are aware the analytical solution of this model appears for the first time here.

The simplified deterministic model provides valuable insights when it comes to analyzing the full model with stochastic outcomes and multidimensional investment. For simplicity we work with a model that has just three quality states: a firm can have higher, equal, or lower quality relative to its rival. After demonstrating existence we use the numerical algorithm of Pakes and McGuire (1994) and Ericson and Pakes (1995) to find and analyze the set of MPE's for this full model. Multiple equilibria are present – we draw on our earlier analytic results to characterize the region of the parameter space in which they occur, and argue for uniqueness in other regions. Within this region of uniqueness we examine the effects on innovation of changes in two parameters that help determine industry 'competitiveness': 'transport costs' (i.e., the degree of heterogeneity in consumer tastes) and a 'switching cost', which measures a firm's cost of moving locations from one end of the Hotelling line to the other.

As a benchmark case we initially consider situations where the costs of differentiation are so high that firms in equilibrium do not switch locations at all – they just invest in improving quality. Since this case is a discrete-time analogue to the type of quality-ladder dynamic duopoly studied in Aghion et al. (2001, 2005) and Budd et al. (1993), an initial contribution of the paper is thus to study how changing the time frame for firms' investment decisions affects the innovation-competition relationship. We introduce changes in the degree of market power by allowing consumers' transport costs to vary. In contrast with the pattern suggested by Aghion et al. (2001, 2005), we find that with discrete time the innovation-competition relationship can be U-shaped rather than inverse-U shaped. The explanation for the difference rests in part on the response of quality leaders to lower transport costs. With discrete time this response may be positive, whereas in the continuous-time version of the model it would be zero. A positive investment response from quality leaders helps offset any tendency for an industry to spend more time in asymmetric-quality states as the degree of competition increases.

We next move to cases where switching costs are low enough to encourage firms to engage in some shifting of locations. In such cases a quality laggard may seek to increase horizontal differentiation by 'running away from' the leader, while the leader may or may not seek to 'give chase', depending on the size of its quality advantage. For some parameter values there will be an absorbing location state (or states), where the firms in the long run end up permanently co-located or permanently separated, while for other parameter values there will be no absorbing location state and the equilibrium will see continual shifting between locations in the long run. All else equal, the latter type of outcome is likely to be observed when consumers' taste heterogeneity (transport costs) is at intermediate levels.

For if preferences are very heterogeneous, then firms prefer to exploit their potential market power by choosing separate locations. If there is low taste heterogeneity then a quality laggard has little incentive for horizontal differentiation, and in the long run the firms end up co-located (i.e., selling the same product).

With respect to the industry's long-run innovation rate, we again find that the probability of an advance in the technological frontier is not necessarily inverse-U shaped in consumer transport costs. In fact this relationship may be quasi-convex rather than quasi-concave. The intuition is fairly straightforward. With *fixed* locations average quality investment is higher when firms are co-located than when they are separated. When transport costs are high firms are permanently separated in the long run; when transport costs are low the firms are permanently co-located. Hence as transport costs fall from high to low (i.e., as the industry becomes 'more competitive') average quality investment must eventually rise from the low level associated with permanent separation to the high level associated with permanent co-location. Furthermore in the permanent-separation region average industry investment may initially decrease as transport costs fall – an overall U-shape for the innovation relationship would then result.

A different pattern emerges w.r.t. the effect of changes in the switching cost. The key here is the industry's initial long-run equilibrium position for high levels of switching cost. This could be permanent co-location or permanent separation, depending on the relative size of the profits accruing to the leader of the quality race. With fixed locations the long-run average rate of innovation is again lower under permanent separation than under permanent co-location. As switching costs fall, the industry moves into situations of continual location switching. Starting from, for example, an initial position of permanent separation, this means that the industry starts to visit the 'other' state of co-location more often, and overall average investment therefore rises. The opposite is true if the initial position is one of permanent co-location. Thus lower switching costs could cause innovation to either rise or fall overall, depending on taste heterogeneity and on the profits accruing to the leader in the quality race. All else equal, when these profits are relatively high the industry will initially be in a state of permanent co-location, and low switching costs would then cause average quality investment to fall.

Our discrete-time framework is a contrast with the continuous-time approach used in, e.g., Aghion et al. (2001, 2005) and Budd (1993). A continuous-time model may be simpler to analyze in some respects. Certain cases of such models imply asynchronous investment by quality laggards and leaders. That is, in states of asymmetric quality only one firm (the laggard) invests at any given instant in equilibrium. This property considerably simplifies the equilibrium analysis. However in practice there are no doubt many industries where firms continue to invest in innovation, even when leading the quality race. In part this may reflect firms' uncertainty about the status of rivals' R&D efforts, or standard cycles for the introduction of new products (e.g., annual industry conventions). Both such explanations call for a model where firms operate under conditions of discrete time.

The structure of our model is closely related to that used in Langohr (2003). However the questions of interest are quite different. Langohr also applies the computational methodology of Pakes and McGuire (1994) to analyze a dynamic duopoly. In her model each firm may in any period sell a high-end product, a low-end product, or both. Each firm may also in each period invest in a single-dimensional quantity called 'capability', increases in which raise average consumer tastes for its product range. To guarantee smoothness there is also a firm- and product-specific logit error in consumer tastes for the available goods.

Langohr's aim is to find conditions under which the firms' product ranges exhibit 'convergence', in the sense of similar product ranges, i.e., both firms offer both types of good, rather than respectively specializing into high- and low-end products. Hence her focus is rather different to ours, which is on the interaction between the dynamics of horizontal and vertical differentiation for single-product firms. In Langohr's model there is no explicit measure of horizontal differentiation, so it would not be possible to infer from her results any of our conclusions about firms' strategic shifts in architecture, and their interactions with quality investment.

The next section analyzes the model with a fixed quality differential and deterministic location transitions. Section 3.3 introduces the stochastic model. The computation of equilibria in this stochastic model is discussed in section 3.4, and its equilibrium value functions are examined in section 3.5. In section 3.6 we consider innovation in the stochastic model with fixed locations, and then move on to consider innovation in the full model with both vertical and horizontal differentiation in section 3.7. Section **??** contains concluding remarks and suggestions for extensions.

Our analysis currently focuses on the impacts of parameter changes on the innovation rate, or on the industry's average investment in quality improvement. Implicit in this focus is the assumption that our duopoly model could be embedded in a growth model (as in, e.g., Aghion and Howitt (1992), Grossman and Helpman (1991)), in which case there would be a monotonic mapping from innovation rates to growth rates. Note that our introduction of horizontal differentiation would add an extra dimension to a full analysis of welfare effects in this model, a task which we leave for future work.

3.2 A deterministic model

Consider a standard Hotelling duopoly where the two firms 1 and 2 may locate at either end (possibly the same end) of the unit line. Let $a_i \in \{0, 1\}$, $i = 1, 2$, denote a firm's location, or product architecture. Consumers are uniformly distributed between 0 and 1 and have unit demands and quadratic transport costs with parameter α. Each firm i produces a single non-durable good at constant marginal cost of c_i. For simplicity let $c_i = 0$.

The firms' goods each have an idiosyncratic quality q_i. A consumer located at $\tau \in [0, 1]$ who purchases a good of quality q at price p from a firm located at $a \in \{0, 1\}$ realizes net utility of:

$$u = -\alpha(a - \tau)^2 + q - p\,.$$

Consumers are fully informed about prices and qualities in the current period and purchase from the firm offering the highest net utility. We will assume throughout that $q_i > 3\alpha$ for both firms $i = 1, 2$. This ensures that the market is always covered in equilibrium.[5]

Denote the difference between the qualities of the firms as $\delta \equiv q_2 - q_1$. Consider the Nash equilibrium of the one-shot simultaneous price-setting game in which firms' locations and product qualities are treated as exogenous. Let π_i denote the equilibrium profits of firm $i = 1, 2$. If the firms are located apart, with, e.g., $a_1 = 0$, $a_2 = 1$, we have:

a. if $\delta > 3\alpha$, $\pi_1 = 0$, $\pi_2 = \delta - \alpha$.

b. if $-3\alpha \leq \delta \leq 3\alpha$,

$$\pi_1 = \frac{\alpha}{2}[1 - (\frac{\delta}{3\alpha})]^2 \ , \ \pi_2 = \frac{\alpha}{2}[1 + (\frac{\delta}{3\alpha})]^2 \ .$$

c. if $\delta <$ -3α, $\pi_1 = -\delta - \alpha$, $\pi_2 = 0$.

These are also the equilibrium payoffs when $a_1 = 1$, $a_2 = 0$. Cases (a) and (c) here represent situations where one firm's quality advantage is sufficient to allow it to sell to the whole market in equilibrium.

When firms are located together ($a_1 = a_2 = 0$ or $a_1 = a_2 = 1$) we have the Bertrand outcome for homogeneous goods. Then non-zero equilibrium profits are only realized when a firm has a strict quality superiority over its rival:

a. if $\delta > 0$, $\pi_1 = 0$, $\pi_2 = \delta$.

b. if $\delta = 0$, $\pi_1 = \pi_2 = 0$.

c. if $\delta < 0$, $\pi_1 = -\delta$, $\pi_2 = 0$.

A key observation here is that, whether firms' locations are separate or coincident, equilibrium profits depend only on the difference in product qualities δ, not directly on the quality

[5]We assume that any consumer who is indifferent between purchasing from either firm randomizes 50-50 between the two.

levels themselves. This observation holds because the market is by assumption covered in equilibrium, and it allows us to adopt a 'quality ladder' approach in what follows. Furthermore it can be seen that firm 2 (resp. firm 1) gets a higher payoff in the 'co-located' equilibrium than in the 'separate' equilibrium if and only if $\delta > 3\alpha(2-\sqrt{3}) \approx 0.804\alpha$ (resp. $\delta < -3\alpha(2-\sqrt{3})$).

Note also that a firm which advances its quality differential from $\delta < 0$ to $\delta = 0$ in states where $s = 1$ receives incremental profits of:

$$K_L = \begin{cases} \frac{\alpha}{2} & \text{if } |\delta| > 3\alpha, \text{ or} \\ \frac{\delta}{6}\left[2 - \frac{\delta}{3\alpha}\right] & \text{if } |\delta| \leq 3\alpha. \end{cases}$$

A firm which advances its quality differential from $\delta = 0$ to $\delta > 0$ receives incremental profits of:

$$K_H = \begin{cases} \delta - \frac{3\alpha}{2} & \text{if } |\delta| > 3\alpha, \text{ or} \\ \frac{\delta}{6}\left[2 + \frac{\delta}{3\alpha}\right] & \text{if } |\delta| \leq 3\alpha. \end{cases}$$

For all α and δ the incremental profits of the quality leader are decreasing in α, while those of the quality laggard are increasing in α, i.e., $dK_H/d\alpha < 0$ and $dK_L/d\alpha > 0$. These signs will help determine the effect on quality investment of changes in α in the full model to be analyzed below. (If $s = 0$ then stage-game profits are independent of α.)

Turn then to dynamic extensions of this Hotelling game in which firms may invest to switch locations and improve their future qualities. Time is discrete and the horizon is infinite. To represent firm i's location, quality, etc. in period t we write $a_i(t)$, $q_i(t)$, $t = 0, 1, 2, \ldots$. Initial values are given by $a_i(0)$, $q_i(0)$. We initially define the state of the industry at the beginning of any period t to be a vector $\omega(t) = ((a_1(t), a_2(t)), (q_1(t), q_2(t)))$ showing the current configuration of firms' respective locations and quality levels. Let $\pi_i(\omega(t))$ denote firm i's stage-game payoffs when the state is $\omega(t)$.

We start with a model of deterministic transitions, in which firms' qualities are exogenously fixed throughout, and their investment decisions only affect their locations.

Future profits are discounted at the rate β. The sequence of moves within each period is as follows:

a. Firms observe the current state $\omega(t)$ and simultaneously decide whether or not to 'switch', i.e., to move to the other location.

b. The switching decisions are implemented and any costs of these decisions are incurred.

c. Conditional on their new locations firms play the Hotelling stage game and receive their profits.

d. Play moves on to the next period, with state $\omega(t+1)$ determined by the location choices implemented in (b).

A firm incurs a fixed cost of $T > 0$ each time that it switches location.

We restrict attention to Markov-perfect equilibria, in which firms' strategies depend only on the current state of the industry, and in which the state only contains 'payoff-relevant' information. A firm's strategy comprises an investment decision, i.e., a state-conditional choice of whether or not to switch locations. Firms' quality levels may be dropped from the state vector $\omega(t)$ since they are held fixed in the present model – let their difference be $\delta(t) = \delta$ for all t. Furthermore it will be apparent from the preceding discussion of the stage game that, since the distribution of consumers along the real line is symmetric, the difference between the location configurations $(a_1 = 0, a_2 = 0)$ and $(a_1 = 1, a_2 = 1)$ is merely one of labelling: both yield the same stage-game payoffs. The same is true of the configurations $(a_1 = 0, a_2 = 1)$ and $(a_1 = 1, a_2 = 0)$. In keeping with the 'payoff-relevant' restriction on state variables in the MPE concept (Maskin and Tirole (1988a,b)), we therefore ignore the distinction between these pairs of location configurations. A state $\omega(t)$ is now redefined to be an element of the set $S = \{0, 1\}$, where $\omega(t) = 0$ represents 'firms are together' and $\omega(t) = 1$ represents 'firms are apart'.

Firm i's strategy is then a function $h_i : S \to \{0,1\}$, $i = 1,2$, where $h_i(\omega) = 0$ means 'stay' and $h_i(\omega) = 1$ means 'switch'. We restrict attention to MPE's in pure strategies, ruling out randomizations between switching and not switching. Where convenient we will use ω to denote the current state and ω' to denote next-period states. Given a rival's strategy h_j, the Bellman equation representing firm i's optimization problem in this game is:

$$V_i(\omega) = \max_{h_i(\omega)\in\{0,1\}} \big(-Th_i(\omega) + \pi_i(\omega') + \beta V_i(\omega') \big) \ , \tag{3.1}$$

where, with $\oplus$ representing binary addition, $\omega' \equiv \omega \oplus h_i(\omega) \oplus h_j(\omega)$. An equilibrium may be represented as a pair of value functions $(V_1(.), V_2(.))$ and a pair of policy functions $(h_1(.), h_2(.))$ such that for $i = 1,2$, $V_i(.)$ solves (3.1) given $h_j(.)$, $j \neq i$, and such that

$$h_i(\omega) \in \arg\max_{h_i\in\{0,1\}} \{ -Th_i + \pi_i(\omega') + \beta V_i(\omega') \} \ , \tag{3.2}$$

for all ω, given $h_j(.)$, where $\omega' \equiv \omega \oplus h_i \oplus h_j(\omega)$.

For any given parameter values it is possible to solve for a pure-strategy equilibrium of this game analytically (if one exists). Suppose for example that $\delta \geq 3\alpha$. Then firm 1 earns zero profits no matter what the state, and any MPE must have $h_1(0) = h_1(1) = 0$, $V_1(0) = V_1(1) = 0$. Any other strategy would have firm 1 incurring some switching costs without any benefit. Firm 2's best response then depends on the parameters $(\delta, \alpha, T, \beta)$: if $\alpha < T(1-\beta)$ we have $h_2(0) = h_2(1) = 0$, $V_2(0) = \delta/(1-\beta)$, and $V_2(1) = (\delta-\alpha)/(1-\beta)$ in equilibrium, while if $\alpha > T(1-\beta)$ we have $h_2(0) = 0$, $h_2(1) = 1$, $V_2(0) = \delta/(1-\beta)$, and $V_2(1) = -T + (\delta/(1-\beta))$. The intuition for these best responses is straightforward: given that firm 1 is not going anywhere, firm 2 finds it optimal to switch in the 'apart' state ($\omega = 1$) if and only if the NPV of the extra stage-game profit, $\alpha/(1-\beta)$, exceeds the one-off switching cost T. (Recall that with $\delta \geq 3\alpha$ firm 2 earns a stage-game profit of δ when $\omega = 0$ and profit of $\delta - \alpha$ when $\omega = 1$.)

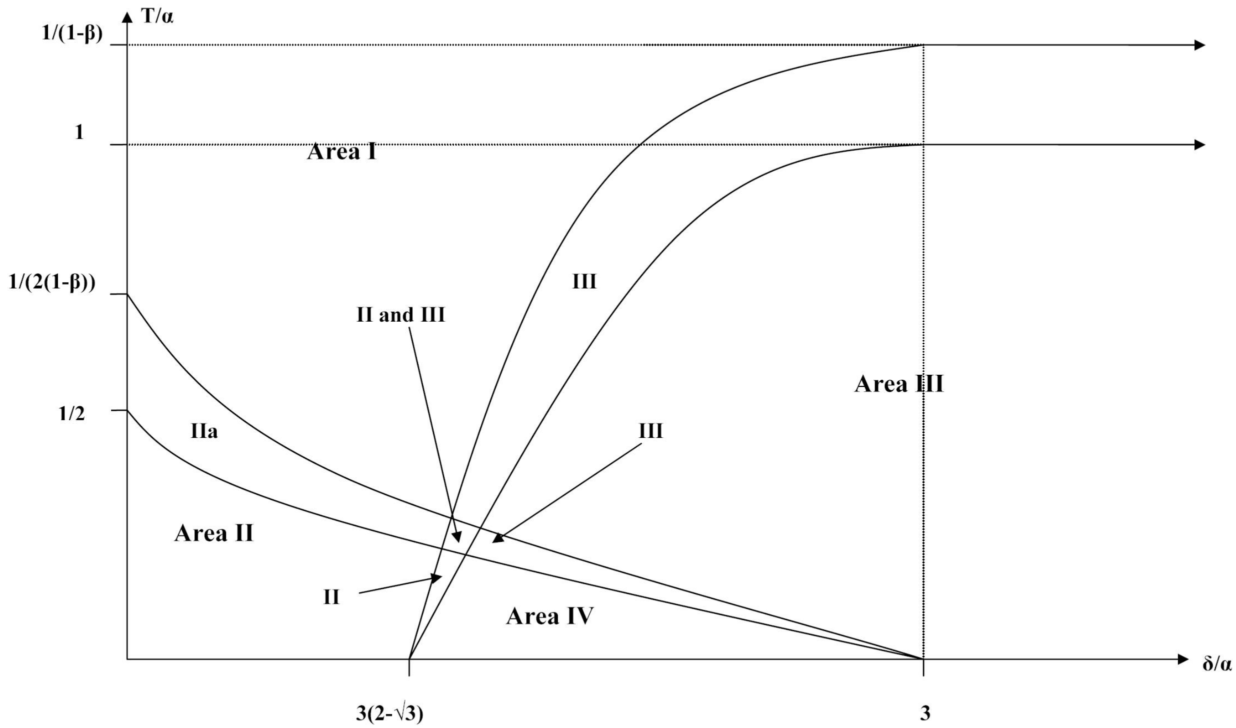
Deriving the pure-strategy equilibria of this game for all possible parameter values is a laborious task. For brevity we therefore omit the details and summarize the results of this effort in figure 1, which characterizes the type of equilibrium outcome according to the location in R_+^2 of the point $(\delta/\alpha, T/\alpha)$. (Note that the figure restricts attention to $\delta \geq 0$: by symmetry a mirror-image picture would apply in the case of $\delta < 0$.) The figure divides this space up into four areas. Three of these represent different types of 'steady states', while the fourth represents the case of no equilibrium in pure strategies.

Take firstly the case of switching costs T that are high relative to α. Area I in the figure indicates that if T exceeds $\alpha/(1-\beta)$ then there will be no movement at all in equilibrium, regardless of the difference in product qualities, i.e., $h_1(0) = h_2(0) = h_1(1) = h_2(1) = 0$. Clearly the equilibrium steady-state location then just depends on the industry's initial state $\omega(0)$. The same type of outcome also obtains for moderately high levels of switching cost, as long as the difference in product qualities is not too high. When δ is larger and T is at these moderate levels we move into area III, where the only equilibrium steady state is $\omega(t) = 0$: the firms end up together. In this region, for given values of T and α, the higher value of δ makes it worthwhile for firm 2 to switch locations to join firm 1 when the state is $\omega = 1$. As long as T/α is not too low, the low-quality producer, firm 1, will not respond to this move by itself switching locations (i.e., it will not 'run away').

If $\delta \geq 3\alpha$ then firm 1 will in fact stay put no matter what the values of T and α, since in those cases it earns zero profit in all states. But if $\delta < 3\alpha$ then firm 1 can make some profit in the apart states. Whether it finds it optimal to run away then depends on T and α – it will set $h_1(0) = 1$ if switching costs T are small enough, or if the consumer transport cost α is high enough (which would give firm 1 more local monopoly power). These cases are represented in the figure by areas II and IV.

In area IV there is no equilibrium in pure strategies. Firm 1 always finds it optimal to switch when $\omega = 0$ and firm 2 is not switching. But for the parameter values in this area firm 2 *ceteris paribus* prefers to be at the same location as firm 1. (Note that $\delta > 3(2-\sqrt{3})\alpha$

Figure 1: Equilibrium outcomes in the model with fixed δ and deterministic location transitions



Area I: outcome is s=0 or 1 (no switching in equilibrium)
Area II: outcome is s=1 (firms end up separate)

Area III: outcome is s=0 (firms end up colocated)
Area IV: no equilibrium

in this region.) Furthermore since T is not too high firm 2 is willing to pay the cost of 'following' firm 1, i.e., if $\omega = 0$ and firm 1 is switching then firm 2 also wants to switch, in which case it would be better for firm 1 to stay put and we have a contradiction. Although there is no pure-strategy MPE in this region there would be a mixed-strategy MPE where both firms randomize between switching and not switching.

Finally in area II we have the case of low switching costs (or high α) and a low quality differential. Except where it overlaps area III, the equilibrium steady-state outcome in this area is always $\omega = 1$: the firms stay apart. The reasoning is simple: for obvious reasons the low-quality firm always prefers to be apart, and since δ is relatively low this is also true of the high-quality producer. Furthermore since switching costs are relatively low one or other of the firms will always be willing to shift away when $\omega = 0$.

A twist in this case is that this steady-state outcome can be arrived at via multiple equilibria. Two obvious such equilibria have: (i) $(h_1(0) = 0, h_1(1) = 0, h_2(0) = 1, h_2(1) = 0)$, and (ii) $(h_1(0) = 1, h_1(1) = 0, h_2(0) = 0, h_2(1) = 0)$. That is, an equilibrium involves only one firm moving away, but either firm may be the mover. There are also two other Pareto-dominated equilibria, in which firms needlessly swap places when they are already apart; these have: (iii) $(h_1(0) = 0, h_1(1) = 1, h_2(0) = 1, h_2(1) = 1)$, and $(h_1(0) = 1, h_1(1) = 1, h_2(0) = 0, h_2(1) = 1)$. In both cases firms could improve their payoffs by agreeing to stay put when they are already apart, but given that my rival fails to do this, it will be optimal for me to respond by also switching, in order to stay away from him.[6]

Area II is the only region of multiple equilibria – outside this area the MPE is unique if it exists. In the region where areas II and III overlap we have not just multiple equilibria, but also multiple steady states (from a given initial state). In particular the equilibrium steady state could be either $\omega = 0$ or $\omega = 1$. The latter arises with the strategies $(h_1(0) =$

[6]Note that multiple equilibria of this type do not exist throughout area II. Close to the origin all four equilibria exist. But in part of the area marked IIa (in particular, the part where δ is highest), the unique equilibrium has $(h_1(0) = 1, h_1(1) = 0, h_2(0) = 0, h_2(1) = 0)$.

$1, h_1(1) = 0, h_2(0) = 0, h_2(1) = 0$): firm 1 moves away when the firms are together, and (since δ is not high enough relative to T/α) firm 2 does not find it optimal to follow him. The former arises with the strategies $(h_1(0) = 0, h_1(1) = 0, h_2(0) = 0, h_2(1) = 1)$. Here firm 1 no longer finds it optimal to move away because his time apart would only last one period before his rival moved to join him. And since firm 1 is not moving when the firms are together it is optimal for firm 2 to switch when they are apart. Note the potential role of commitment in resolving the multiplicity of steady states here – if firm 1 could commit to always switching in co-located states he could deter firm 2 from ever attempting to join him. Similarly if firm 2 had the commitment power he could deter firm 1 from ever attempting to move away.

We have dwelt in some detail on the equilibrium outcomes in this deterministic model for two reasons. First, it is analytically tractable and we can therefore be certain of which types of equilibria occur where. Second, the intuition behind this case carries over to the model with stochastic transitions, where we will see a similar pattern of equilibrium outcomes to that apparent from figure 1. For reasons of tractability that model must be solved numerically, however the intuitions drawn from figure 1 will help reinforce our confidence in the thoroughness of our computational approach.

3.3 A stochastic model

We now allow firms to invest not just in changes in their architecture, but also in quality improvements. Firms' potential locations are still restricted to the endpoints of $[0, 1]$. Their quality improvements will be subject to a form of interdependence that reflects the possibility of spillovers arising from firms' investment successes. Moreover we will allow the results of investment (in either architecture changes or quality improvements) to have stochastic outcomes. Since these complications render the model analytically intractable we turn to the computational algorithm of Pakes and McGuire (1994) in order to solve for the MPE.

As previously let $s(t) \in S = \{0, 1\}$ denote the industry's differentiation state at the start of period t. Once again we abstract away from the payoff-irrelevant labels of the two possible locations, and just focus on whether firms are together $(s(t) = 0)$ or apart $(s(t) = 1)$. Let Δ be a finite subset of the real line representing the set of possible quality differences $\delta(t) \equiv q_2(t) - q_1(t)$. For simplicity we will henceforth restrict Δ to just three elements: $\Delta \equiv \{-k, 0, +k\}$ for some $k > 0$. The value k represents the size of a single step in the transitions of the quality differential δ. At the beginning of any period t the industry's quality differential can be in any one of three states: firm 1 is one step ahead $(\delta(t) = -k)$, the firms are of equal quality $(\delta(t) = 0)$, or firm 2 is one step ahead $(\delta(t) = k)$. By varying k we can vary the value of success in the firms' quality-investment competition.

Let the vector $\omega(t) \equiv (s(t), \delta(t)) \subset S \times \Delta$ denote the state of the industry at the beginning of period t. Firms' absolute quality levels $q_i(t)$ are omitted from this vector, since only the difference in qualities matters for the stage-game payoffs. We simply assume that each firm's quality starts at $q_i(0) > 3\alpha$: since there is no depreciation of quality the market will then be covered in any equilibrium of the Hotelling stage game.

Within each period the sequence of moves is as follows:

a. at the beginning of period t firms observe the current state $\omega(t)$ and simultaneously choose levels of investment in architecture shifts and quality improvement.

b. stage-game payoffs are realized given the current state.

c. the investment plans chosen in (a) are implemented, their outcomes are realized, and the costs of the plans are incurred.

d. play moves on to the next period, with the new state $\omega(t + 1)$ determined by the realizations of the investment plans implemented in (c).

Note that the order of moves here is slightly different to that in the deterministic model of the previous section. Previously investment plans (changes in architecture) were implemented prior to the stage-game play in each period. Here they are implemented after the stage game.

State transitions in this model will be stochastic, because the outcome of each firm's investment activity is subject to some uncertainty. Consider first investment in architecture changes (i.e., location changes), which we will sometimes term 'switching' investment for brevity. Let $h_i(t) \geq 0$ denote the dollar value of investment of this type by firm i in period t. With probability

$$A(h_i(t)) \equiv \frac{\exp(-\gamma)h_i(t)}{1 + \exp(-\gamma)h_i(t)} \tag{3.3}$$

the investment is successful, where γ is a cost parameter. Success in switching investment means that the firm changes its location from $a_i(t) = 0$ to $a_i(t+1) = 1$, or vice versa. As γ increases the dollar cost $h_i(t)$ of implementing any given probability of switching success also increases. The function $A(.)$ is strictly concave. Of course a firm that does not wish to change locations can ensure this outcome by setting $h_i(t) = 0$.

Whether switching success translates into a change in the industry's differentiation state $s(t)$ depends on the outcome of the other firm's switching investment. If both succeed in switching then there is no change in s; if one switches and the other does not then s changes from 0 to 1, or vice versa. Since all investments are chosen simultaneously each firm thus chooses its own $h_i(t)$ taking as given the switching probability $A(h_j(t))$ implied by the rival's investment. We assume that the stochastic outcomes of each firm's switching investments are independent. Where the differentiation state is represented s for this period and s' for next period, we thus have:

- $s' = s$ with probability $\theta_{(h_1,h_2)}(s, s') \equiv (1 - A(h_1))(1 - A(h_2)) + A(h_1)A(h_2)$.
- $s' = 1 - s$ with probability $\theta_{(h_1,h_2)}(s, s') \equiv A(h_1)(1 - A(h_2)) + (1 - A(h_1))A(h_2)$.

With respect to investment in quality improvements we allow firms to condition their respective investment levels on the outcome of their switching efforts. That is, we imagine that each firm simultaneously develops two plans for quality investment – one to be implemented in the event that the firm switches locations, and one to be implemented if the firm does not switch. Represent the dollar values of these two conditional investment levels as $v_i^s(t)$ and $v_i^f(t)$, respectively. The firm need only incur one of these investment costs, depending on whether or not its switching investment is successful.[7] Note also that these investments may only be conditioned on the own-firm's switching success. Since the outcome of the rival's switching efforts is not observed until after all investment plans have been determined, quality investments may not be conditioned on any current-period (deterministic or stochastic) choice of the rival. This automatically precludes conditioning quality investment on this period's realized change in the differentiation state, i.e., on $s(t+1)$.

We make the stochastic outcomes of firms' quality investments interdependent by placing an upper bound on the absolute value $|\delta|$ of the inter-firm quality differential. Thinking of a firm's quality improvement as occurring in discrete steps of k units per step, we restrict the quality differential to never exceed one step: thus $|\delta| = 0$ or k. It is assumed that a firm whose quality advantage has reached the upper bound k cannot realize any further quality improvements without automatically conferring the same increment on the quality of its rival. Our motivation for this restriction is that the low-quality firm can reverse engineer the products of the high-quality firm, and thus avoid falling too far behind. However there are limits to the speed and usefulness of this reverse engineering, and therefore quality differences do not evaporate instantaneously and may persist over time.

When a firm has achieved a quality advantage of $\delta = k$ it may nevertheless continue to invest in quality improvements, not to extend its advantage, but to prevent its lower-quality rival from catching up.[8] Formally, let $\tilde{q}_i(t)$ be a random variable representing po-

[7]Thus for example a firm might establish two separate teams of quality planners, each of which prepares the contracts (with suppliers etc.) that will be needed in the event of a particular switching outcome. The firm only actually incurs the costs of quality investment when it sees its switching outcome, after which it signs the contracts produced by one of the teams of planners.

[8]This is an important difference with continuous-time models. If successful quality improvement depends only on current investment outlays (and not on an accumulated stock of 'know-how') then in a continuous-time model a firm which has achieved the maximum permissible quality advantage will not invest anything further until the laggard starts to catch up.

tential increments in firm i's quality at time t arising *from its own investment efforts*. For a firm that has invested an amount $v_i(t) \geq 0$ in quality improvement this increment is equal to $+k$ with probability

$$Q(v_i(t)) \equiv \frac{\exp(-\phi)v_i(t)}{1 + \exp(-\phi)v_i(t)} \,, \tag{3.4}$$

and is equal to zero otherwise. Here ϕ parameterizes the cost of implementing any given probability of success in quality improvement. We assume that, conditional on firms' actual quality investments $v_i(t)$ and $v_j(t)$, the random variables $\tilde{q}_i(t)$ and $\tilde{q}_j(t)$ are independent (and also i.i.d. over time).

If a firm i is not a quality laggard, i.e., if $q_i(t) \geq q_j(t)$, then its realized quality increment in period t conditional on $v_i(t)$ will be $q_i(t+1) - q_i(t) = \tilde{q}_i(t)$. However if the firm is a low-quality producer, $q_i(t) < q_j(t)$, then its quality can be improved not just by its own efforts but, alternatively, by spillovers from the innovations of its rival. That is, $q_i(t+1) - q_i(t) = \tilde{q}_i(t) + \not\Vdash(\tilde{q}_i(t) = 0)[\tilde{q}_j(t)]$, $i \neq j$, where $\not\Vdash$ represents the indicator function. Note that this specification allows the laggard i to benefit from positive realizations of either $\tilde{q}_i$ or $\tilde{q}_j$, but not both.

This specification of the evolution of firms' quality levels results in the following transition probabilities for the state variable $\delta(t)$. Let $\theta_{(v_1,v_2)}(\delta, \delta')$ denote the probability of transition from δ to δ', given quality investments v_1 and v_2. We then have

$$\begin{aligned}
\theta_{(v_1,v_2)}(k,k) &= 1 - Q(v_1)(1 - Q(v_2)) \\
\theta_{(v_1,v_2)}(k,0) &= Q(v_1)(1 - Q(v_2)) \,,
\end{aligned}$$

and

$$\begin{aligned}
\theta_{(v_1,v_2)}(0,k) &= Q(v_2)(1 - Q(v_1)) \\
\theta_{(v_1,v_2)}(0,-k) &= Q(v_1)(1 - Q(v_2)) \\
\theta_{(v_1,v_2)}(0,0) &= 1 - \theta_{(v_1,v_2)}(0,k) - \theta_{(v_1,v_2)}(0,-k) \,,
\end{aligned}$$

and

$$\begin{aligned}\theta_{(v_1,v_2)}(-k,0) &= Q(v_2)(1-Q(v_1))\\ \theta_{(v_1,v_2)}(-k,-k) &= 1-Q(v_2)(1-Q(v_1))\ ,\end{aligned}$$

and $\theta_{(v_1,v_2)}(k,-k)=\theta_{(v_1,v_2)}(-k,k)=0$, because firms cannot improve their quality by more than one step per period.

A firm's strategy is now a vector $u_i(\omega)\equiv(h_i(\omega),v_i^f(\omega),v_i^s(\omega))$ of (non-negative) real-valued functions of the payoff-relevant state $\omega(t)\equiv(s(t),\delta(t))$. Firm i's best response to a rival's strategy $u_j(\omega)$ will be a policy function generated by the firm's Bellman equation:

$$V_i(\omega)=\max_{u_i=(h_i,v_i^f,v_i^s)\in R_+^3}\left\{\pi_i(\omega)-h_i-A(h_i)v_i^s-(1-A(h_i))v_i^f+\beta\mathrm{E}_{\omega'}[V_i(\omega')\mid\omega,u_i,u_j]\right\}\ . \tag{3.5}$$

Using the transition probabilities developed above the expectation on the RHS of (3.5) can be written as:

$$\begin{aligned}&\mathrm{E}_{\omega'}[V_i(\omega')\mid\omega,u_i,u_j]=\\ &\sum_{\delta'\in\{-k,0,+k\}}\Big\{V_i(1-s,\delta')[A(h_1)(1-A(h_2))\theta_{(v_1^s,v_2^f)}(\delta,\delta')+(1-A(h_1))A(h_2)\theta_{(v_1^f,v_2^s)}(\delta,\delta')]\\ &+V_i(s,\delta')[A(h_1)A(h_2)\theta_{(v_1^s,v_2^s)}(\delta,\delta')+(1-A(h_1))(1-A(h_2))\theta_{(v_1^f,v_2^f)}(\delta,\delta')]\Big\}\ ,\end{aligned} \tag{3.6}$$

where $\omega=(s,\delta)$.

We seek an MPE in symmetric pure strategies. That is, we are looking for a pair of policy functions $(u_1(\omega),u_2(\omega))$ and a pair of value functions $(V_1(\omega),V_2(\omega))$ such that, for $i,j=1,2$, $i\neq j$: (i) $V_i(\omega)$ solves (3.5) given $u_j(\omega)$, (ii) given $V_i(\omega)$ and $u_j(\omega)$, the function $u_i(\omega)$ solves the RHS of (3.5) for all ω, and (iii) for all $\omega=(s,\delta)$, $u_2(s,\delta)=u_1(s,-\delta)$. This latter condition is the symmetry requirement: firm 2's chosen action when the differentiation state is s and his quality advantage is $\delta = q_2 - q_1$, should be the same as firm 1's chosen action given s and the same quality advantage $-\delta = q_1 - q_2$.

Existence of MPE's in dynamic oligopoly models is extensively discussed in Doraszelski and Satterthwaite (2005). These authors exploit the arguments of Whitt (1980) to show the existence of an equilibrium in symmetric pure strategies in the framework introduced by Ericson and Pakes (1995). Their existence proof has two key steps. First, it is necessary to show that the fixed point V_i^* that solves the Bellman equation (3.5) is continuous in the rival's strategy u_j. Second, for any given rival's strategy u_j, firm i's best response u_i^* needs to satisfy a unique investment choice (UIC) criterion, i.e., it needs to be single valued. Without policy functions satisfying UIC we cannot be sure that the equilibrium will be in pure strategies, and hence that it will be computationally tractable.

For completeness we include below an abbreviated version of the Whitt-Doraszelski-Satterthwaite existence argument that is tailored to our simple framework. In this proof the first step mentioned above is handled by showing that the contraction operating on V_i in (3.5) preserves continuity in u_j. To see the second step – that any policy function generated by V_i^* satisfies UIC – take for example the case where the current state $\omega = (s, \delta)$ has $s = 0$ and $\delta = 0$. Define $dV_i(s, \delta) \equiv V_i(s, \delta) - V_i(s, \delta - k)$ for $\delta = 0$ or $+k$, and for brevity let $A_i \equiv A(h_i)$, $Q_i^s \equiv Q(v_i^s)$, $Q_i^f \equiv Q(v_i^f)$, for $i = 1, 2$. It can then be seen that for $i = 2$ the expectation in (3.6) can be rewritten as:

$$\begin{aligned}(1 - A_2)&\left\{Q_2^f \mathrm{E}_{a_1', \tilde{q}_1'}[dV_2(s', \delta') \mid s = 0, \delta = 0, a_2' = a_2, \tilde{q}_2 = k, u_1]\right\} \\ &+ A_2\left\{Q_2^s \mathrm{E}_{a_1', \tilde{q}_1'}[dV_2(s', \delta') \mid s = 0, \delta = 0, a_2' \neq a_2, \tilde{q}_2 = k, u_1]\right\} + R\,, \quad (3.7)\end{aligned}$$

where R represents terms that are all linear in A_2, with coefficients on A_2 that depend on V_2^*, u_1, the current state and the exogenous parameters (and not on v_2^f or v_2^s). The first expectation in (3.7), for example, represents firm 2's expected gain from an investment success ($\tilde{q}_2 = k$) next period given that $\omega = (0, 0)$, that firm 2 does not switch locations, and given the rival's investments represented in u_1. (Recall that $Q_2^f \equiv Q(v_2^f)$ represents firm

2's probability of success in raising its own quality, given that it does not switch locations.) This expectation is taken over the rival's stochastic switching and quality outcomes $(a_1', \tilde{q}_1')$. Similarly the second expectation in (3.7) represents the expected next-period gain from a quality improvement conditional on firm 2 changing its location.

While (3.7) (when multiplied by the discount factor) shows the unconditional expected benefits from raising the probabilities of quality improvement, the unconditional expected costs of this effort will be $\mathrm{E}[v_2 \mid u_2] \equiv A_2 v_2^s + (1 - A_2) v_2^f$ (see (3.5)). Since the quality investments can be conditioned on the firm's switching outcome, their optimal levels v_2^{s*} and v_2^{f*} will in fact be independent of A_2, the firm's chosen switching probability, given a value function V_2^* that solves (3.5). Instead they will depend only on the current state, the rival's strategy, and the exogenous factors. Furthermore by the strict concavity of $A(.)$ and $Q(.)$ these optimal investment levels will be unique:

$$\begin{aligned} v_2^{f*} &= \max\left\{0, \frac{\sqrt{\beta x_2^f(\omega, u_1^*, V_2^*)\exp(-\phi)} - 1}{\exp(-\phi)}\right\} \\ v_2^{s*} &= \max\left\{0, \frac{\sqrt{\beta x_2^s(\omega, u_1^*, V_2^*)\exp(-\phi)} - 1}{\exp(-\phi)}\right\} \end{aligned}$$

where

$$x_2^f(\omega, u_1, V_2) = \mathrm{E}_{a_1', \tilde{q}_1'}[dV_2(s', \delta') \mid \omega, a_2' = a_2, \tilde{q}_2 = k, u_1]$$

and

$$x_2^s(\omega, u_1, V_2) = \mathrm{E}_{a_1', \tilde{q}_1'}[dV_2(s', \delta') \mid \omega, a_2' \neq a_2, \tilde{q}_2 = k, u_1]\,.$$

A caveat here is that v_2^{s*}, the quality investment conditional on switching, need not be unique if $A_2 = 0$. For then there is no positive probability of firm 2 switching and so any level of v_2^s could be optimal. We will impose uniqueness in this case by assuming that the optimal v_2^s is still equal to the level that would be optimal if A_2 were to be strictly positive. Following Doraszelski and Satterthwaite (2005) we justify this assumption by imagining that there may be some small probability that players make errors in their choices of switching investment.

Given the optimal quality investments v_2^{f*} and v_2^{s*} we can derive the implied optimized success probabilities Q_2^{f*} and Q_2^{s*} and substitute these into (3.7). Substituting the resulting expression into the RHS of (3.5) results in:

$$\max_{h_2 \in R_+} \Big\{ \pi_2(\omega) - h_2 - A(h_2) v_2^{s*} - (1 - A(h_2)) v_2^{f*} \\ + \beta [(1 - A(h_2)) Q_2^{f*} x_2^f(\omega, u_1^*, V_2^*) + A(h_2) Q_2^{s*} x_2^s(\omega, u_1^*, V_2^*) + R(A(h_2), \omega, u_1^*, V_2^*)] \Big\} .$$

The maximand here is linear in $A(h_2)$, with a coefficient on $A(h_2)$ that represents the firm's expected net benefit from changing locations, and which may be either positive or negative. If it is positive then the optimal switching investment h_2^* is unique by the strict concavity of $A(.)$. If it is negative then the optimal switching investment h_2^* is zero. It follows that at state $\omega = (0, 0)$ the optimal choice of (h_i, v_i^f, v_i^s) generated for firm 2 from V_2^* is unique. We can repeat essentially the same arguments at any other state to show that the policy function generated by V_2^* must satisfy the UIC criterion. Clearly the symmetry in the two firm's problems implies that the same can be said about firm 1.

Given that any solution to (3.5) must generate a single-valued policy function, we can proceed to our existence result.

Proposition 1 *(Whitt (1980), Doraszelski and Satterthwaite (2005)) An MPE in symmetric pure strategies exists for all γ, ϕ and all $k, \beta > 0$.*

As will be seen below, in our model uniqueness of equilibrium need not hold at all parameter values, although we will use our numerical results to argue that it holds at some parameter values.

It will be apparent from the preceding discussion that one of our motivations for conditioning quality investment on the outcome of a firm's switching investment is to ensure that UIC is satisfied. If instead we were to require firms to invest a fixed amount

$v_i^s = v_i^f = v_i$ regardless of their switching outcome it would be considerably harder to determine analytically whether the policy function for (3.5) is single valued. In particular this fixed amount would now depend on the switching probability $A(h_2)$ (since we would have $Q_2^f = Q_2^s$ in (3.7)). The maximand on the RHS of (3.5) will then in general no longer be strictly quasiconcave in h_2. Langohr (2003) encounters this problem and resorts to numerical techniques (a grid search) in order to verify UIC. Since such numerical tests represent a considerable computational complication the practical advantages of our specification should be apparent.

An alternative approach would be an incomplete-information model in which a firm's chosen location in any period is a response to a privately observed draw from an idiosyncratic distribution of location-specific fixed costs. This would be a variation on Doraszelski and Satterthwaite's (2005) incomplete-information specification of the Ericson and Pakes (1995) model of dynamic oligopoly with endogenous entry and exit. To see how such approaches could be applied here, suppose that at the beginning of each period each firm i, $i = 1, 2$, privately observes the value of a random variable ϵ_i^l for each location $l = 0, 1$. These errors, or 'types', represent the firm's fixed costs of operating at l in that period and are i.i.d. across firms, locations and periods. Strategies for each firm are then simultaneous state- and type-conditional deterministic choices of location and quality investment in each period. Since neither firm observes the other's type the rival's choice of investment in any period becomes a random variable from the point of view of the own firm.

Existence of equilibrium in this model can be demonstrated using the arguments of Doraszelski and Satterthwaite, and the approach to computation of equilibrium would be similar to that explained below. For present purposes we stick with our complete-information approach since it provides somewhat clearer distinctions between the different types of equilibria. In particular when the distribution of the errors in the incomplete-information approach has infinite support each location will be chosen with strictly positive probability by each firm in every state in every period. As will be seen below this is not true in our preferred approach, where zeroes often appear in firms' equilibrium switching probabilties. Insofar as they provide a clear indication of firms' differentiation strategies, such zero probabilities make the interpretation of equilibrium outcomes easier, and facilitate comparisons with the deterministic model discussed in section 3.2.

3.4 Computation of equilibria

We rely on the algorithm of Pakes and McGuire (1994) to solve for the symmetric pure-strategy MPE in our model numerically. Starting with initial guesses V_2^0 and u_2^0 for the equilibrium value and policy functions for firm 2, at the k-th iteration of the algorithm we:

a. Convert u_2^k into its symmetric analogue for firm 1, u_1^k.

b. Plug V_2^k and u_1^k into the maximand on the RHS of (3.5) and solve for firm 2's optimal investments (h_2, v_2^f, v_2^s) at each state ω.

c. Evaluate the maximand at the optimal investments to get $V_2^{k+1}(\omega)$ at each ω, set u_2^{k+1} equal to the optimal investments, and return to (a).

Iteration is terminated when $\max(\|u_2^{k+1} - u_2^k\|, \|V_2^{k+1} - V_2^k\|)$ falls below some prespecified tolerance, where $\|.\|$ represents the sup norm over states ω.[9]

Given the small numbers of firms and states in our model the computational burden of this algorithm is usually not excessive. When coded in Matlab 7 the algorithm typically takes around 45 seconds to calculate an equilibrium to a tolerance of 10^{-9} on a PC with a 2.7GHz processor. If an MA(4) dampening procedure is used (see Judd (1998)) the algorithm is fairly stable and does not often fail to converge. An exception is when switching costs are extremely low and taste heterogeneity is high. In such cases the algorithm often

[9] Note that, as discussed in Pakes and McGuire, the algorithm does not completely solve firm 2's Bellman equation before updating the rival's strategy. Instead iteration through the operator defined in (3.5) and updating of the rival's strategy proceed together.

has difficulty converging and we were not always able to find an equilibrium. Our numerical analysis therefore avoids this region of the parameter space, effectively assuming that firms' switching costs exceed some (very low) minimal level.

Figure 3.2 shows the various types of ergodic set that arise in equilibrium at different parameter values. In this figure the parameter determining the difficulty of quality investment is fixed at $\phi = 0$ and taste heterogeneity α is set at 1. Shown on the horizontal axis are different values for k, the size of the step up when quality investment is successful. The parameter γ determining the difficulty of switching investment is on the vertical axis. Higher values of this parameter correspond to more costly switching investment.

Recall that in drawing figure 1 we were able to normalize the quantities on the horizontal and vertical axes by α. That figure showed steady-state outcomes for different combinations of δ/α, the quality differential per unit of transport costs, and T/α, the switching cost per unit of transport cost. Whether a similar normalization is possible in the present more general framework is an open question. In other words we do not yet know whether α has direct effects on equilibrium outcomes other than through its effects on simple indices such as (for example) $\gamma/\log(\alpha)$ and k/α. Nevertheless, for ease of comparison with figure 1, in figure 3.2 we 'normalize' k by α on the horizontal axis. Figure 3.3 shows an equivalent picture with $-\log(\alpha)$ on the vertical axes (and γ set to zero). Since the patterns of equilibrium outcomes are identical between the two figures it would appear that some type of normalization may be possible. The effects of *ceteris paribus* changes in α are still a valid object of study, regardless of whether or not such a normalization is possible. All else equal, when α rises the industry passes along a path such as that indicated by the curve superimposed onto figure 3.3 (moving from the northeast to the southwest).

The ergodic sets arising from the equilibrium transition probabilities are classified into four types according to the nature of the absorbing differentiation state (or states). For example if $h_i(.) = 0$ at all ω s.t. $s = 0$ then co-location of firms is an absorbing differentiation state – once together, firms will never separate along the equilibrium path.

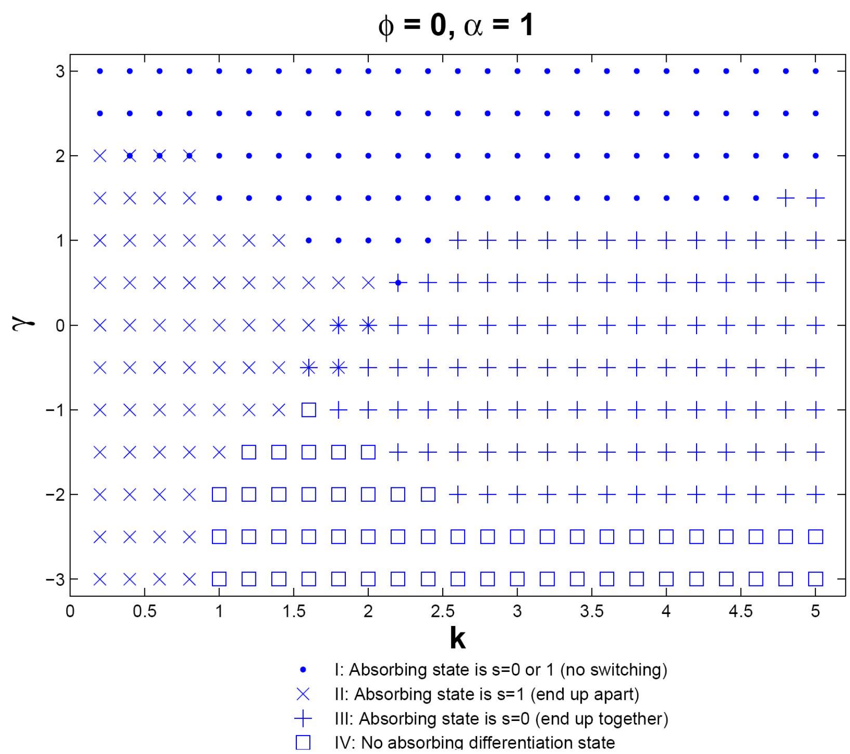


Figure 3.2: Equilibrium outcomes in the model with endogenous δ and stochastic transitions

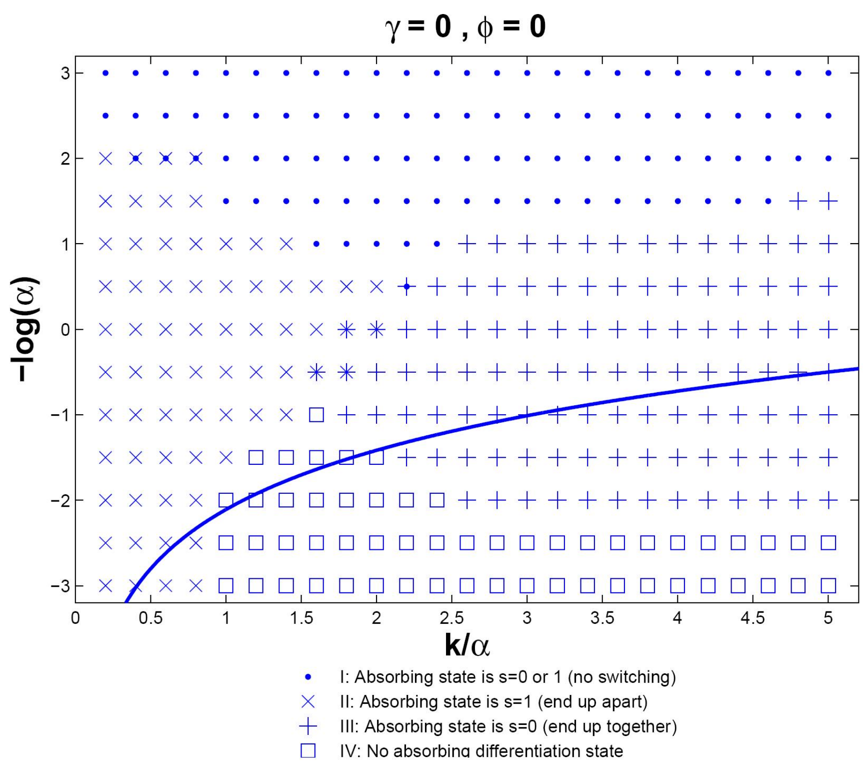


Figure 3.3: Equilibrium outcomes in the model with endogenous δ and stochastic transitions – effects of changes in α

If in addition $h_i(.) = 0$ at all ω s.t. $s = 1$ then there is no switching at all in equilibrium, in which case separate locations is also an absorbing differentiation state. On the other hand we take cases where $h_i(.) \neq 0$ for some state $\hat{\omega}$ s.t. $s = d$, where $d = 0$ or 1, to imply that d is not an absorbing differentiation state. Strictly speaking for this to be true we should also require that in equilibrium there be a positive probability of reaching the state $\hat{\omega}$ from any other state $\tilde{\omega}$ where $s = d$ and $h_i(.) = 0$ (so that the industry doesn't get 'stuck' at $\tilde{\omega}$). In practice this extra condition is almost always satisfied over the parameter ranges considered below, because at least one firm invests in quality improvement in each state. We will explicitly point out any cases where the extra condition fails due to zeroes in quality investment.[10]

Figure 3.2 shows a pattern of equilibrium types consistent with that of figure 1. At the top of the figure is a region of no movement in equilibrium (area I – indicated by bold dots). This arises because of the high value of γ in this region: switching locations is expensive, so no-one ever moves. On the right-hand side of the figure, crosses (area III) represent equilibria where co-location is the only absorbing differentiation state. Here switching costs are moderate and k, the 'prize' in terms of quality differentials, is relatively large. Moderate switching costs and a high k encourage a firm that is a quality leader to move close to the laggard when $s = 1$. On the other hand when k is high the laggard has less to gain from moving away – therefore he does not respond in kind by switching, and as a result the firms will always end up together.

If switching costs fall to a low enough level then the industry moves from the co-location outcome into area IV, represented by the squares, where there is no absorbing differentiation state. That is, conditional on $s = 0$ or 1, along any equilibrium path there is always a strictly positive probability of reaching the other differentiation state $1 - s$. For example when $s = 1$ the quality leader wants to switch so as to join the laggard and

[10]The condition may not be satisfied if quality investment is very costly, i.e., ϕ is high. Then we may find both firms setting $v_i^f = 0$ and $h_i(.) = 0$ in some state $\tilde{\omega}$, in which case $\tilde{\omega}$ is an absorbing state in itself. Such values of ϕ generally fall outside the parameter ranges considered below.

increase its stage-game payoffs. But if he succeeds in doing this, thereby changing the differentiation state to $s = 0$, the laggard may then want to move away. The result (once shifts in the quality differential are also factored in) is endless cycling through all possible states. Finally when the quality differential k is low enough a quality leader will lose his incentive to co-locate with his rival, since his stage-game payoffs are higher when they are separate. Then both players want to differentiate, regardless of the state, and $s = 1$ is the only absorbing differentiation state. In the figure such outcomes are represented by x symbols – area II.

It will be apparent that for each of these four regions there is a corresponding region in figure 1 with a similar intuition. Furthermore in both figures we observe cases of 'multiple steady states', i.e., multiple types of equilibrium outcome for a given set of parameters. In figure 2 these are represented by the superposition of two of the aforementioned symbols. For example when $k = 2$, $\gamma = 0$ there are two types of outcome – one in which co-location is the only absorbing state and one in which separation is the only absorbing state. Note the analogy with the overlap between areas II and III in figure 1. Moreover in figure 2 we see new areas of overlap not observed in figure 1. When $k = 2.2$ and $\gamma = 0.5$, for example, there is an equilibrium with no switching at all and an equilibrium where the firms will always end up together. Similarly there is also an overlap between the area of 'no movement' and the area 'end up apart'.

By analogy with figure 1 we should also expect to observe some cases of multiple equilibria of the same type, particularly in area II. In that region both firms want to locate apart. There are two ways of achieving this: either the quality leader can move away, or the laggard can move away. Both behaviors may be consistent with equilibrium (but not with the same equilibrium). Such cases of multiple equilibria are in fact observed for parameters in this area of figure 3.2 (although these multiples are not indicated on the figure for brevity).

The model of figure 3.2 also brings with it a new instance of multiple equilibria associated with coordination failures in quality investment. Consider for example the case of equal-quality firms located apart ($\omega = (1,0)$). Suppose for simplicity that they do not invest in switching. One equilibrium may be for neither firm to invest in quality improvement, in which case $\omega = (1,0)$ is an absorbing state – both firms 'cruise', incurring no investment costs and earning stage-game payoffs of $\alpha/2$ per period. There may also be an equilibrium where both firms invest in quality improvement in this state – given that one's rival invests something it will be optimal for the own firm to do likewise. Such cases of multiple equilibria in quality investment are frequently observed in the present model when k is low and ϕ, the difficulty of quality improvement, is high. A low value of k means that a firm's payoff to (temporarily) winning the quality investment race is low – then firms will be willing to settle for the 'cruise' outcome of no quality investment. Similarly this outcome is also relatively attractive when quality upgrades are expensive – as ϕ falls this equilibrium will disappear because firms start to 'defect' from the zero-investment strategy profile.

In summary, we find cases of multiple equilibria for one of three reasons:

a. near a boundary between regions representing two different classes of ergodic sets.

b. in area II, because k is low, the firms want to separate, and either the leader or the laggard may be the one to shift away.

c. where k is low and ϕ is high, because of failures in the coordination of investment between firms of equal quality.

As long as we are restricted by the complexity of the problem to numerical techniques we can never claim with 100% certainty to have found all equilibria at given parameter values without *a priori* information on the number of such equilibria to be found. Since we had difficulty computing equilibria at high α and low switching costs, we restrict $\alpha < \exp(2)$, $\gamma > -3$ in our subsequent analysis. Although lacking a fully analytic characterization of equilibria in our model, we believe it likely, given these restrictions, that the above three classifications cover all multiple equilibria in this model. Our confidence stems from a numerical search over a grid of 12,675 points in $(k, \gamma, \phi, \alpha)$ space, using eleven different starting values per point. This search did not throw up any instances of multiple equilibria other than those covered by (a), (b) and (c).

In view of the above comments we henceforth focus on parameter values such that: the quality step k is not too small relative to α ($k/\alpha > 2.2$) and the cost of quality investment is not too high ($\phi \leq 1$). For $\gamma \in (-3, 3]$ and $\log(\alpha) \in [-3, 2)$ the numerical searches found no instances of multiple equilibria over $(k/\alpha, \phi)$ in this range. When for illustrative purposes our analysis strays outside this region of the parameter space we will explicitly note that fact. To get an idea of the range of environments covered by these parameters, note that if $\alpha = 1$, $k = 2.2$, and $\phi \leq 1$ then the cost to a firm of inducing a 50% probability of success in quality improvement (i.e., $\Pr(\tilde{q}_i = k) = 0.5$) would be no more than 1.24 times the firm's maximum stage-game payoff. For the same parameters a value of γ in $(-3, 3]$ means that the cost to the firm of inducing a 50% probability of switching success is between 0.02 and 9.1 times the maximum stage-game payoff.

3.5 Equilibrium value functions

Figure 3.4 shows the equilibrium value function V_2^* for various parameter values. This is represented by two curves, each of which shows V_2^* for the three possible quality differentials $\delta \in \{-k, 0, +k\}$. One curve represents states where the firms are apart and the other represents states where they are together. The plots in the top, middle and bottom rows respectively show the effects of a higher payoff to a quality leader (raising k), of more costly switching investment (raising γ), and of more costly quality investment (raising ϕ). In each row the centre panel shows the value functions for the parameters $k = 3.2$, $\gamma = -1$, $\phi = -1$, $\alpha = 1$.

Increasing the differential k clearly raises the expected profits of a quality leader, through its effect on the leader's stage-game payoffs. In the figure this change also has a positive effect on the value function at states of equal or inferior quality ($\delta = 0$ or $-k$).

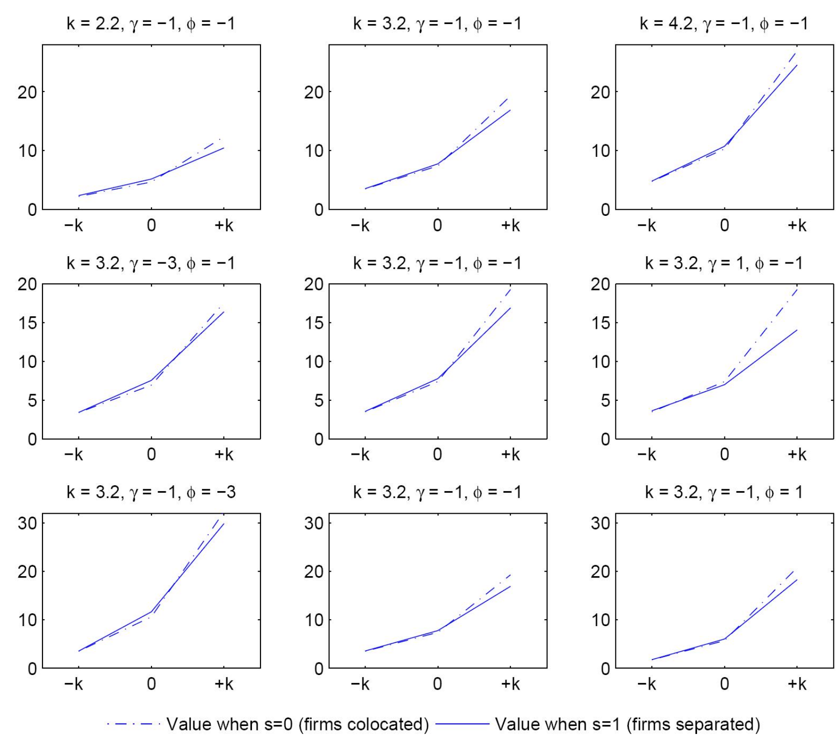


Figure 3.4: Equilibrium value functions, conditional on the horizontal differentiation state s, as functions of the quality differential δ, for $\alpha = 1$

Since raising k does not directly increase the stage-game payoffs of firms in these states (e.g., if $\omega = (0, 0)$ then $\pi_1 = \pi_2 = 0$ always), this is an indirect effect. Since investment outcomes are stochastic, quality leadership is impermanent – eventually firms' present quality rankings will be reversed and the erstwhile laggard will become the leader. The expectation of this eventual ranking reversal causes payoffs in all states to increase in k, at least for the parameters shown in the diagram.

The middle row of figure 3.4 shows how more costly switching drives a wedge between payoffs in the co-located and separate states. Note in particular that while the value for a firm in the 'best' state of $\omega = (0, +k)$ rises with γ, the value for a firm in the second-best state of $\omega = (1, +k)$ ends up lower at $\gamma = 1$ than at $\gamma = -3$. For a firm in the former state the reason for the rise in V_2^* is fairly clear: as switching becomes more expensive it is harder for the firm's lower-quality rival to 'escape' from co-location. Hence the firm can expect to enjoy the high profits from its current position for longer (barring any reversal in quality rankings).

For a firm in the state $(1, +k)$ there are (at least) two effects working in opposite directions. All else equal, the rise in $V_2^*(0, +k)$ referred to above should also pull up $V_2^*(1, +k)$, since a firm at $s = 1$ has some possibility of switching in the next period. On the other hand as γ increases such a switch becomes more costly, which will tend to reduce $V_2^*(1, +k)$. The figure indicates that over some parameter ranges the latter effect can dominate the former.

Relative to the deterministic model discussed earlier, this divergence between the value functions in separate and co-located states can lead to reversals in firms' preferences among locations. Consider in particular the states (0,0) and (1,0). In the deterministic model (with δ fixed) both firms prefer to be in the latter state – in state (0,0) they earn zero stage profits until one of them moves away. With stochastic quality upgrades the opposite ranking of states may hold. This can be seen in the rightmost panel of the middle row in figure 3.4, where it is seen that for relatively high switching costs we have $V_2^*(0, 0) > V_2^*(1, 0)$, i.e., equal-quality firms have higher expected profits when they are together than when they are apart. The reason is that with costly switching the gains from winning the quality investment race are considerably higher at $s = 0$ than at $s = 1$ – the possibility of realizing these gains induces a preference for co-location. (Note here that the value functions are convex in δ, i.e., firms are risk lovers with respect to quality improvements.)

Effects of more costly quality investment are shown in the bottom row of figure 3.4. Firms in the equal- or lower-quality states ($\delta = 0$ or $-k$) see expected profits fall as ϕ rises. This is intuitively reasonable: in those states the cost of winning the quality race is increasing without any compensating increase in stage-game payoffs. For firms in the high-quality states ($\delta = +k$), the effect is non-monotonic. Expected profits may initially decline in ϕ – compare $\phi = -3$ with $\phi = -1$ – and later rise. The reason for this is that the dominant position of a quality leader is most at risk when the cost of quality investment ϕ is at intermediate levels. When ϕ is high it is costly for either firm to raise its own quality; when ϕ is low it is easy for *both* firms to raise quality – in either case it will be difficult for the laggard to catch up. Hence a quality leader enjoys a relatively protected position when ϕ is either high or low.

3.6 Innovation with fixed locations

As an application of our framework we study the effects of changes in the competitive environment on the industry's investment in innovation. In our model innovation can affect welfare in two dimensions – through the horizontal differentiation of products and through improvements in their quality. For the present we focus on investment in that latter, vertical, dimension of product improvement, in order to compare the predictions of our framework with those of, e.g., Aghion et al. (2001, 2005). Those authors studied the effects of competition in dynamic duopolies where firms invest in cost reduction. Under certain specifications of demand and the investment technology, firms' instantaneous equilibrium payoffs in those models depend only on the gap between firms' costs, not on the absolute levels of costs. Similarly in our model firms' stage-game payoffs depend only on the gap between firms' qualities, not on their absolute levels – in that sense the comparative statics and dynamics of any cost-reduction model should map into those of the same model reformulated as one of quality improvement.

In this section we study quality investment with fixed locations, i.e., when switching is so expensive that in equilibrium a firm's switching investment $h_i(\omega)$ is set at zero for all states ω. We vary α, consumers' transport costs, to see how changes in the heterogeneity of consumer tastes affect innovation in the quality dimension. Note that the sign of the effect of α on the *levels* of stage-game profits depends on whether a firm gets to make all the sales when it is the quality leader.[11] However for the determination of quality investment it is the profit *increments* from successful innovation that ultimately matter. These are the quantities K_L and K_H defined in section 3.2 (for states where $s = 1$), and they are respectively monotonic increasing and decreasing in α, for all δ/α. That is, when firms are separated, a laggard benefits more from quality improvement when α is higher, and a leader benefits less. This fact plays a central role in determining the response of innovation to changes in the heterogeneity of consumer preferences.

Figure 3.5 shows $Q(v_i^{f*})$, a firm's equilibrium state-conditional probability of success in quality investment, given $s = 1$. In this figure the quality investment cost ϕ is set equal to -1.5; figure 3.6 shows the same curves for the higher level of investment cost $\phi = 1$. Three curves are shown in each figure, for the three states $(s = 1, \delta = -k), (s = 1, \delta = 0), (s = 1, \delta = +k))$, as functions of α, at four different values of the innovation size k. Setting the switching cost γ equal to five means that there is no switching in equilibrium: if the firms start out separated they stay separated forever. Thus in calculating the probability of investment success for these figures we do not need to take account of a firm's switching-conditional investment v_i^{s*}.

[11] That is, from the profit expression in section 3.2 we see that, if $|\delta/\alpha| < 3$ then both firms make positive sales in any state where $s = 1$, and in such states we have $\partial\pi_i/\partial\alpha > 0$ for $i = 1, 2$. On the other hand if $|\delta/\alpha| > 3$ then in such states the high-quality firm makes all the sales and its profit decreases in α.

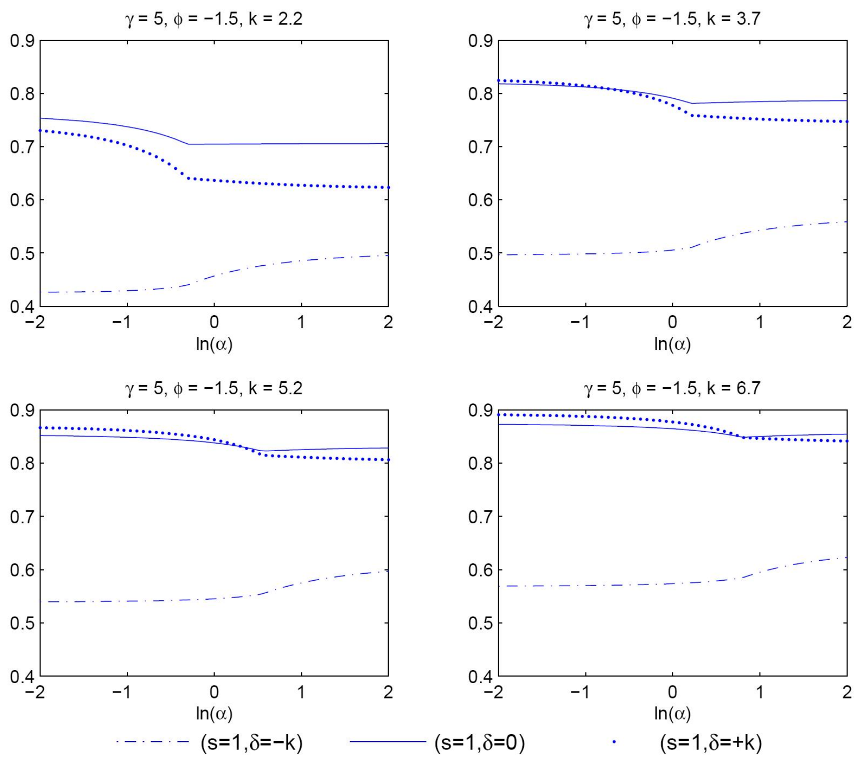


Figure 3.5: Equilibrium state-conditional probabilities of success in quality investment with fixed locations, as functions of α, given $s = 1$, for $\phi = -1.5$

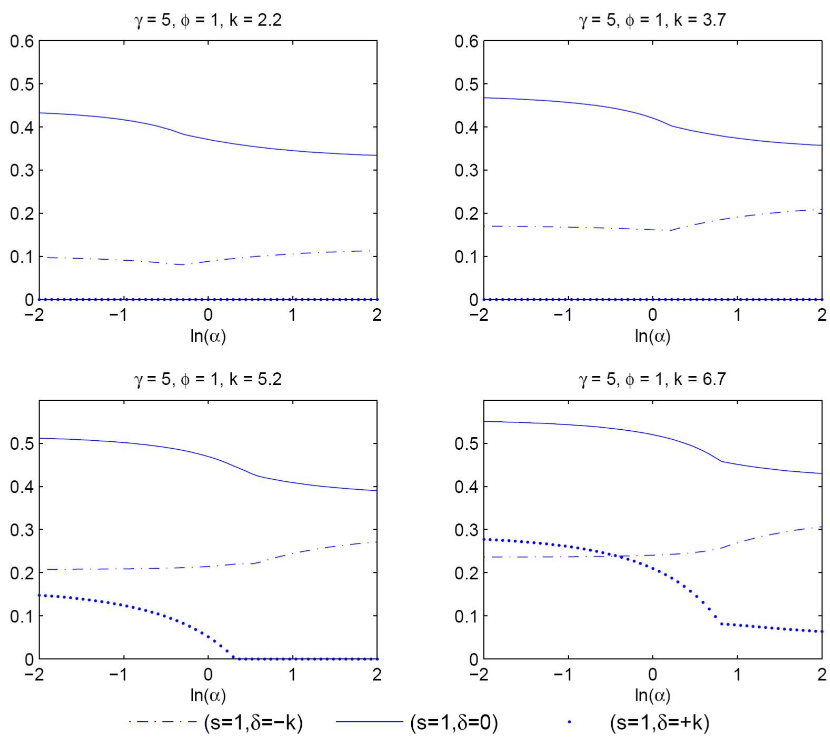


Figure 3.6: Equilibrium state-conditional probabilities of success in quality investment with fixed locations, as functions of α, given $s = 1$, for $\phi = 1$

Both figures show that the average level of quality investment by firms in the 'neck-and-neck' equal-quality state (to use the terminology of Aghion et al. (2001)) is higher than average investment by firms in asymmetric-quality states. (Compare the height of the solid curve with the average height of the two other curves.) This pattern is consistent with results from continuous-time models in the existing literature. It may not be immediately obvious that the pattern should carry over to the discrete-time model considered here, because in the present case equilibrium quality investment by a firm which has achieved the maximum quality advantage may be non-zero, in contrast with continuous-time models. Furthermore it is clear from figure 3.4 that in this model a quality leader has fairly strong incentives to maintain that standing, since the value functions are convex in the quality differential, being steepest when a firm moves from $\delta = 0$ to $\delta = +k$ or vice versa. Nevertheless average *industry* quality investment is still seen to be lower in the asymmetric-quality states, because of the low investment by the laggard. In turn this allows the leader to slacken his investment somewhat, leading to a lower average overall.

Figures 3.5 and 3.6 show that a leader's quality investment declines with rises in α, the heterogeneity in tastes. To see the reason for this consider the effects of α on the profit increments K_H and K_L defined in section 3.2. Any firm bases its quality investment decisions in part on some weighted sum of these increments. The weights will depend on the discount rate, and on the future likelihood of experiencing a transition from $\delta = -k$ to $\delta = 0$, or from $\delta = 0$ to $\delta = +k$, given the rival's investment strategy. Each transition is accompanied by a change in the stage-game profit of K_L, for the former kind, or K_H, for the latter. Naturally a firm that is currently a leader will be *relatively* more likely to experience 'equal-to-high' transitions in future.[12]

To see that in equilibrium a high-quality firm is more likely to experience equal-to-

[12]Here we need not distinguish between transitions up and transitions down, because moves in either direction still involve the same *change* in π, i.e., the same change K_H. That is, an equal-to-high transition arises from state $\delta = 0$ only if my rival's quality investment fails, in which case the payoff to my own success is K_H. A high-to-equal transition arises from state $\delta = +k$ only if my rival's quality investment succeeds, in which case the payoff to my own success is again K_H.

high transitions in future (whether up or down) use a one-stage deviation argument. From $\delta = +k$ this period the state next period will either be unchanged or will go to $\delta = 0$. Symmetric equilibrium behavior from the latter state must imply that the probabilities of going back up to $+k$ or down to $-k$ are equal. Viewed from the present period the overall probability of an equal-to-high transition in future must therefore exceed the probability of low-to-equal transitions, regardless of this period's quality investment.

Consider then the situation of a firm that bases its quality investment decisions on the unweighted average of increments $K_L + K_H$. We have:

$$\frac{d}{d\alpha}(K_L + K_H) = \frac{1}{2} - \frac{3}{2} = -1 < 0 \text{ if } k/\alpha > 3 \tag{3.8}$$

and

$$\frac{d}{d\alpha}(K_L + K_H) = \frac{\delta^2}{18\alpha^2} - \frac{\delta^2}{18\alpha^2} = 0 \text{ if } k/\alpha < 3\,. \tag{3.9}$$

Hence a firm which based its investment decisions on the unweighted average profit increment would have v_i^f non-increasing in α. In practice a quality leader will put relatively more weight on the increments K_H, since they are more likely to be experienced in future. It would then follow that in deciding its optimal quality investment the leader (in part) uses a weighted average of K_L and K_H that is strictly decreasing in α. Therefore its quality investment strictly decreases in α: the intuition is that more consumer heterogeneity always reduces the equal-to-high profit increment K_H, and these increments are relatively over-weighted in the leader's objective function.

Next consider quality investment by firms in quality states $\delta = 0$ or $\delta = -k$. In contrast with investment by a leader, figures 3.5 and 3.6 show that investment by these firms is not monotonic in α. Investment by the laggard 'usually' rises in α, but may decrease in α for low α, low k and a high cost ϕ – see the first panel of figure 3.6. And investment by a neck-and-neck firm usually falls in α, but may rise in α for high α, high k and low ϕ – see the fourth panel of figure 3.5. Note that the break point where the slopes change is always near the value of α such that $k/\alpha = 3$. (In the third panel of figure 3.5, for example, it is around $\exp(0.55) = 1.733 \approx 5.2/3$.) That is, the kinks in these figures are triggered by α rising above $k/3$, i.e., out of the region where consumer tastes are so homogeneous that the quality leader gets all the sales.

For the quality laggard, it would seem that in the region of high α ($\alpha > k/3$), quality investment should always be increasing in α. The reason is the opposite of the explanation just advanced for the case of quality leaders. A laggard will relatively overweight low-to-equal moves among its future possible transitions. These are all accompanied by a change of K_L in stage-game profits, and (3.9) then suggests that if $\alpha > k/3$ the laggard will be basing quality investment on a weighted sum that is strictly increasing in α(i.e., on a sum which has derivative in α of the form $\mu_1(\delta^2/\{18\alpha^2\}) + \mu_2(-\delta^2/\{18\alpha^2\})$, where $\mu_1 > \mu_2$. Thus the probability of investment success should rise in α for $\alpha > k/3$, a prediction which is borne out in the figures.

Note however that changes in α may affect not just the profit increments but also their weights (relative frequencies) in a firm's objective function. Changes in these weights could in some cases have countervailing effects on firms' quality investments. Consider for example the case of low α, low k and high ϕ represented in the first panel of figure 3.6. As noted above, the laggard's quality investment here declines in α. To see the reason for this note that at these parameter values the discounted sum of future industry profits is heavily concentrated in the hands of the current quality leader. Stage-game profits for anyone else are either zero (for a laggard) or very low ($\alpha/2$, for neck-and-neck firms), because α is low. Furthermore in equilibrium the industry has a low tendency to leave asymmetric-quality states, because of the laggard's low quality investment (yielding a success probability of less than 0.1).

As a consequence the main impact of a rise in α on the laggard's objective function at these parameter values is its effect on the expected time it takes to usurp the leader's position in the quality race. This effect is negative for any given level of quality investment by the laggard, because of its effect on investment by equal-quality firms. If his investment succeeds this period, the laggard reaches the neck-and-neck state next period (since the leader invested nothing, as is clear from the figure). From there his chances of achieving quality leadership in the subsequent period would be maximized by a success probability in the neck-and-neck state of $Q(v_i^{f*}) = 0.5$. (That is, the probability $Q(1-Q)$ of a firm advancing to quality leadership from the neck-and-neck state is quasi-concave in Q and maximized at $Q = 0.5$.) The figure indicates that rises in α actually take the equilibrium success probability in the neck-and-neck state further away from this 'ideal' value, reducing it below 0.4. Hence with higher α a laggard who caught up one step with the leader would have to wait longer before making it all the way to the top. Thus (again applying the one-stage deviation principle), a laggard may find that its expected current-period benefit from quality investment falls somewhat as α rises.

For equal-quality firms the next transition could be either equal-to-high or equal-to-low. To determine the effect of α on investment by a firm i in this state, we should consider which of these transitions is most likely, given the rival's strategy and given the cost ϕ, representing the (inverse of the) ease of quality improvement. If ϕ is low then quality improvement is easy and it is relatively more likely that the rival will succeed in improving its quality this period. In that case firm i is investing in quality not so much to get ahead, but just to keep up. That is, its choice of quality investment will be based on a greater likelihood of changes K_L in stage-game profits. Since K_L is increasing in α, we should expect that investment by equal-quality firms may sometimes rise in α if ϕ is low. Figure 3.5 shows this to be true, e.g., for $k/\alpha < 3$ and $k = 6.7$. (Fourth panel – the rise is slight but noticeable.) Note that for $\alpha < k/3$ quality investment still falls in α – the reason is that $dK_H/d\alpha$ jumps up in absolute value when α goes from above $k/3$ to below $k/3$. That is, increases in consumer heterogeneity α have their most negative effects on the leader's incremental profits when he is getting all the sales in the market, and for such α that negative effect is the dominant influence on investment by neck-and-neck firms.

Whereas figures 3.5 and 3.6 show firms' state-conditional success probabilities, figures 3.7 and 3.8 represent the industry's innovation rate under the invariant distribution. That is, they show the equilibrium long-run average probability of an improvement in the industry's frontier technology. This means the probability of an improvement by the leader, in states where $\delta = -k$ or $+k$, or by either firm in states where $\delta = 0$. The invariant distribution weights these probabilities by the amount of time the industry spends in each state over the long run. Because the cost parameter γ is still set high enough to discourage any switching in equilibrium, there are in fact two ergodic sets at these parameter values, one for $s = 0$ and one for $s = 1$. Shown in the figure is the innovation rate for $s = 1$, as a function of α – note that the equivalent figure for $s = 0$ would show no variation in α, because when firms are co-located their stage-game payoffs are independent of α.

The results of Aghion et al. (2001, 2005) are a relevant point of comparison here. Those authors examined the effect of competition on innovation rates in duopoly, measuring competition by an exogenous elasticity of substitution between the products of the two firms. They found that as competition increases (i.e., as the elasticity of substitution rises from zero), the industry's long-run average innovation rate first increases. This is because both firms invest nothing in innovation when there is no competition (i.e., when the elasticity is zero). An increase in competition initially spurs some positive levels of innovation by both firms, leading to an increase in the industry innovation rate. This is the 'escape-competition' effect.

When competition becomes sufficiently intense Aghion et al. find that a countervailing effect sets in. Because innovative activity is most intense when firms are in neck-and-neck competition (as in the present model), the industry tends to leave the neck-and-neck states rather quickly, and over the long run spends more time in states of asymmetric quality (or asymmetric technologies).[13] Therefore the long-run average rate of innovation in the industry starts to decline. This is referred to as a 'composition effect'. The overall picture

[13]Innovation in Aghion et al. is an investment in cost reduction rather than quality improvement. As noted above, it should be possible to recast their models as games of quality improvement.

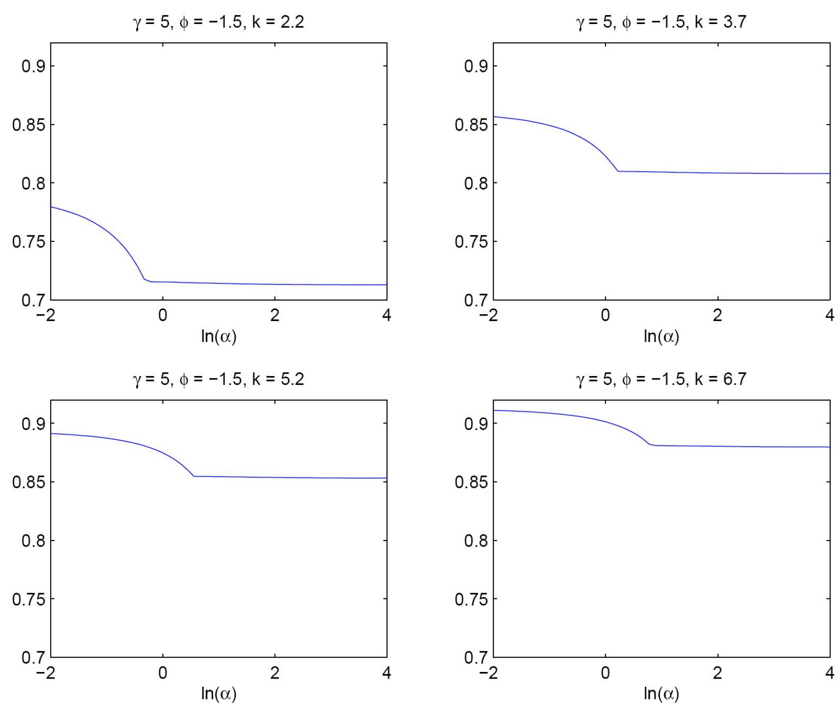


Figure 3.7: Probability of an advance in the frontier technology under the equilibrium invariant distribution, given $s = 1$ and $\phi = -1.5$

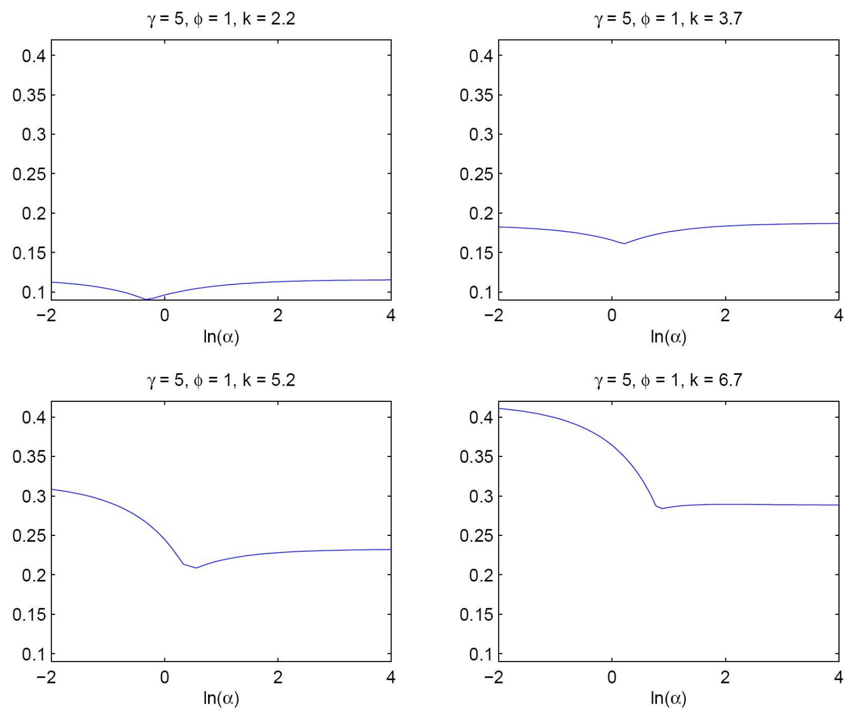


Figure 3.8: Probability of an advance in the frontier technology under the equilibrium invariant distribution, given $s = 1$ and $\phi = 1$

suggested is one of an 'inverse-U' relationship between competition and innovation.

The principal differences between our model and those of Aghion et al. (2001, 2005) are that: (a) we work in discrete time rather than continuous time; and (b) the degree of competition in our model is a function of primitives: the heterogeneity in consumer tastes (α) and the costs of changing locations (γ). Also, Aghion et al. (2001) allow multiple steps in the technological differentiation between firms, i.e., firms can be separated by more than one quality step. As a special case they consider a version of their model in which the leader can be at most one step ahead – this case of their model most closely corresponds to our framework.

Figures 3.7 and 3.8 respectively show innovation rates for cost parameters $\phi = -1.5$ and $\phi = 1$, corresponding to the parameter values in figures 3.5 and 3.6. They reveal a quite different pattern to the inverse U reported by Aghion et al. – if anything quasi-convexity in α seems to hold, rather than quasi-concavity. Note however that quasi-convexity does not hold everywhere. In particular the fourth panel of figure 3.8 indicates that, after first falling then rising in α, the average innovation rate is thereafter again (slightly) declining in α, for high ϕ and high k. To highlight the differences with Aghion et al. we will focus on explaining the clear U-shape indicated in the other panels of that figure.

When consumer transport costs fall from a high level, the innovation rate may decline before rising again. This initial decline reflects the composition effect noted by Aghion et al. As α falls, innovation by laggards decreases – see figures 3.5 and 3.6. Simultaneously innovation by leaders rises. Innovation by neck-and-neck firms may go either way – recall, e.g., the fourth panel of figure 3.5 – but if it falls it does not decline as fast as innovation by laggards. The upshot is that as α falls the industry over the long run spends relatively more time in asymmetric-quality states. In those states the probability of an advance in the frontier technology (i.e., the probability of successful investment by the leader) is relatively low, and as a consequence the overall innovation rate falls as α falls.

Eventually α falls below the threshold level $k/3$, beyond which the firms are competing for (temporary) control of the whole market. Recall that at this point the slope in α of the incremental stage-game profits for an equal-to-high transition jumps up (in absolute value), from $1/2$ (for $\alpha > k/3$) to $3/2$ (for $\alpha < k/3$). Therefore quality investment by neck-and-neck firms begins to rise more strongly, and the same will be true for quality leaders if they are investing positive quantities. Simultaneously quality investment by laggards levels out somewhat, and may begin to increase. This ameliorates the composition effect of lower α, since it tends to send the industry back to the neck-and-neck state sooner. The figures indicate that as a result higher investment by neck-and-neck firms (and possibly also by quality leaders) tends to dominate the weakened composition effect, and the average innovation rate rises overall.

Taken separately an initial decline, and subsequent rise, in the innovation rate as α falls may be interpreted in the terms put forward by Aghion et al. The former, arising from the industry spending less time in neck-and-neck states, is a composition effect, and the latter, arising from an upswing in investment by neck-and-neck firms, is an escape-competition effect. It is striking that in our framework these effects appear in the opposite order to that seen in Aghion et al. That is, as a measure of 'market power' (α) falls, the composition effect dominates first, then the escape-competition effect, rather than the other way around. In summary it would seem that care needs to be exercised in translating the competition effects of Aghion et al. into the effects of primitives such as consumer taste heterogeneity.[14]

[14]Figure 3.7 indicates that the composition effect may be so weak as to be dominated everywhere by the escape-competition effect. Here we are simply discussing one possible pattern in the innovation-competition relationship. Note also that the escape-competition effect may be reinforced by increased investment by quality leaders, although the first and second panels in figure 3.6 indicate that this is not necessary for the effect to dominate.

3.7 Innovation with endogenous locations

Turn now to the effects of parameter changes in situations where switching costs are low enough to induce some movement between locations in equilibrium. We continue to focus on the effects on firms' investment in quality improvement. Figure 3.9 shows state-conditional probabilities of success in quality investment, conditional on $s = 1$, for various values of γ. The top, middle, and bottom panels respectively show success probabilities for the three states $\delta = +k, 0, -k$, as functions of α. Note that the success probabilities are now expectations of the switching-conditional probabilities $Q(v_i^{f*})$ and $Q(v_i^{s*})$, weighted by $A(h_i^*)$, the probability of success in switching.

The key point of this figure is that the endogeneity of location choices only 'matters' when consumer taste heterogeneity is at intermediate levels. In all three panels the solid line shows the probability of investment success when $\gamma = 5$, at which value firms never change locations in equilibrium. At lower levels of γ there may be some movement in equilibrium, depending on the values of the other parameters. Nevertheless the figure shows that for high and low values of α the firms' success probabilities (for states where $s = 1$) all converge, for all quality differentials, and regardless of the value of γ.

To see the reason for this, recall the path indicated for changes in α in figure 3.3. At low α there is effectively little differentiation between locations, and hence there is little reason for firms to seek competitive advantage through horizontal differentiation. In particular quality laggards have little incentive to run away from the leader, and the absorbing state sees both firms co-located. That is, we are in region III in the figure. The low level of α means that a firm's stage-game payoff will be approximately equal across the location states $s = 0$ and $s = 1$, i.e., it is (approximately or exactly) zero for all but the quality leader, who gets (approximately or exactly) δ in each stage. Therefore firms' quality investments conditional on $s = 1$ all converge in γ, for small α.

On the other hand for high α firms have substantial potential market power if they locate separately. That is, when α is high enough we will have $3\alpha(2 - \sqrt{3}) \approx 0.804\alpha > \delta$,

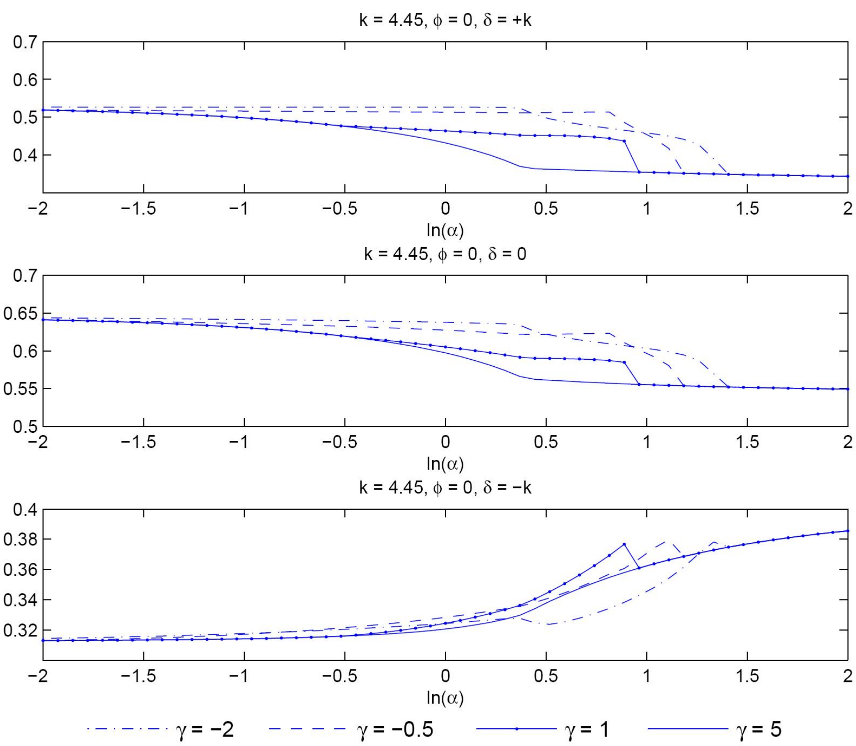


Figure 3.9: Equilibrium state-conditional probabilities of success in quality investment with endogenous locations, as functions of α, given $s = 1$

and therefore firms always prefer separation regardless of their quality advantage. Then the absorbing state sees firms permanently separated. In figure 3.3 we would be in region II. As soon as firms reach this state $s = 1$ they cease all further location switching – hence quality investments conditional on $s = 1$ do not depend on γ.[15] For any γ the value of α at which this situation of permanent separation is reached can be read from figure 3.9 – it is that α at which the probability-of-success curve merges with the solid curve.

For intermediate levels of α there may be no absorbing location state in equilibrium – the industry may pass through region IV in figure 3.3. In this region there is enough heterogeneity in consumer tastes to make it worthwhile for a laggard to run away from a co-located leader. At the same time there is not so much heterogeneity that the leader does not want to give chase. Figure 3.9 indicates that, conditional on separation ($s = 1$), firms' quality investments will then depend on the level of the switching cost. Leaders and neck-and-neck firms seem to invest more when separated than they would if locations were fixed (i.e., compared to investment for $\gamma = 5$), while laggards may invest more or less. Whatever the sign of the effect, figure 3.9 suggests that a model that predicts quality investment by conditioning on α, but without taking into account the endogeneity of firms' locations in the product space, is most likely to 'miss' when consumer taste heterogeneity is at intermediate levels.

Figure 3.10 shows firms' state-conditional probabilities of success in quality investment, as functions of the switching cost γ, given $s = 1$ and at various values of α. Note that at the parameter values shown in this figure we always have $k > 0.804\alpha$, so that location states $s = 0$ maximize the stage-game payoff of a quality leader. A change in γ can then affect these equilibrium success probabilities through two avenues. One is the effect on the proportion of time that the industry spends in the location states $s = 0$ in the long run. Since co-location maximizes the leader's stage-game profits, all else equal he prefers

[15]This statement presumes that the industry does not fall into a Pareto-dominated equilibrium, where firms always prefer separation but in which they continue to engage in needless switching simply because their rival keeps on switching.

longer durations in states $s = 0$. The second avenue is the effect of γ on the probability of the industry moving to co-location from states $s = 1$. As γ rises such moves become more costly.

The overall impact of a change in γ on firms' quality investments depends on the signs of these two types of effect, and on their relative magnitudes. *Ex ante* we cannot say whether an increase in γ will raise or lower the long-run proportion of time that the industry spends in states $s = 0$. On the other hand a large enough such increase must eventually lower the probability of getting to $s = 0$ from states $s = 1$. Hence it would lower the probability with which a quality leader reaches its best state ($s = 0, \delta = +k$), and thus would tend to reduce quality investment by such firms.[16]

Consider then raising γ from low values to high values at each set of parameter values represented in figure 3.10. When γ is low the industry in equilibrium sees continual shifting between locations – there is no absorbing location state. This is true for low γ in all but the first panel shown in figure 3.10. (In that first panel the initial equilibrium has co-location as the absorbing state, i.e., the equilibrium is in region III in figure 3.2.) If on the other hand γ is high then there will be no switching at all in equilibrium, and changes in γ will have no effect on equilibrium investments. Accordingly we see in all four panels that the curves level out for high γ.

Between these two extremes the industry may pass through one of two *unique* absorbing location states. In the first three panels of figure 3.10 that state is co-location. Referring to figure 3.2, note that co-location (area III) can be the unique absorbing state if k/α is large enough, i.e., if α is not too high. We see in those panels that, as γ increases, investment by leaders and neck-and-neck firms initially also rises. This reflects the influence of the first type of effect noted above – since the industry is headed for permanent co-location, the long-run probability of being in states $s = 0$ rises in γ, and leaders and

[16] In an earlier version of this paper we presented results indicating that firms' equilibrium switching probabilities are in fact not everywhere monotonically decreasing in γ, but they will decrease for large enough rises in γ.

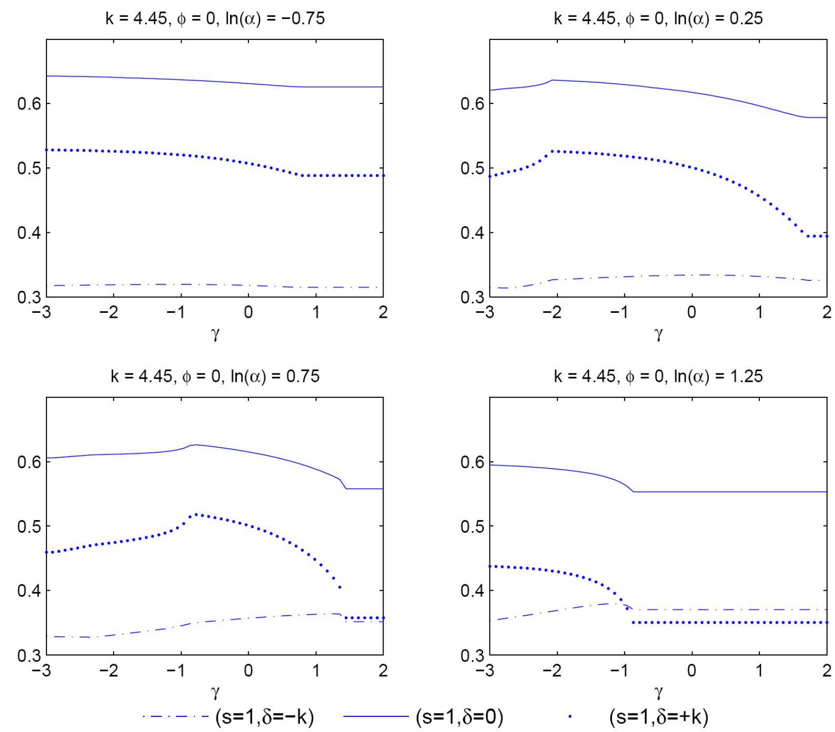


Figure 3.10: Equilibrium state-conditional probabilities of success in quality investment with endogenous locations, as functions of γ, given $s = 1$

neck-and-neck firms invest more in quality improvement. (Laggards also invest more, but the effect seems weaker than for the other firms.) Quality investment by these firms peaks when the absorbing state is reached – thereafter it declines because of the second type of effect noted above: higher γ makes it harder for quality leaders to reach the 'good' state $s = 0$ from $s = 1$.

In the fourth panel a quite different effect is in operation. There k/α is relatively low, and as γ rises the industry is headed for separation as the unique absorbing state (i.e., region II in figure 3.2). As γ rises the long-run probability of co-location then initially decreases, exerting a negative influence on investment by leaders and neck-and-neck firms. This effect is reinforced by the fact that higher γ makes it harder for such firms to get to $s = 0$ from $s = 1$. Thus in the fourth panel we see quality investment by these firms falling up until the point where $s = 1$ is the unique absorbing location state, after which changes in γ have no further influence on quality investment. Investment by laggards is seen to move in the opposite direction (although not monotonically), because as γ rises they are more protected from the possibility of losing all their sales to a co-located quality leader.[17]

It will be apparent from the preceding discussion that the industry's long-run probabilities of co-location and separation play key roles in determining the effects of parameter changes. Note that this effect is not explicitly visible in models such as Aghion et al. (2001) which measure the degree of competition by an exogenous elasticity of substitution. Results in that literature emphasize the role of the composition effect discussed previously – that is, the effect of a parameter change on the proportion of time that an industry spends in neck-and-neck competition. While this effect is also in operation in the present model, it turns out that the 'co-location effect' (or perhaps 'horizontal composition effect') introduced here can be the dominant influence in the determination of an industry's long-run innovation rate.

[17]There is some possibility of multiple equilibria arising at the parameter values in the fourth panel of figure 3.10. The discussion here should thus be taken as a particular example of equilibrium behavior consistent with these parameter values.

To see this consider first figures 3.11 and 3.12, which respectively show for various values of α the industry's long-run probability of co-location, and its long-run innovation rate, as functions of γ. The parameter values under consideration are the same as in figure 3.10. As previously, the long-run innovation rate is defined as the probability of an advance in the technology frontier under the equilibrium invariant distribution.

There is an obvious correspondence between the slopes of the two figures. That is, whether the rate of innovation increases or decreases as γ rises depends on the effect on firms' long-run propensity to be co-located. For relatively low levels of heterogeneity – e.g., $\log(\alpha) = 0.25$ and $\log(\alpha) = 0.75$, corresponding to the second and third panels in figure 3.10 – the industry is headed for zero horizontal differentiation as an absorbing state. Long-run average quality investment by firms that are permanently co-located is higher than in equilibria where they continually shift locations. This is because of the very convex value functions faced by such firms. They receive zero profits in states $\delta = -k$ and $\delta = 0$, but their maximum stage-game profits in states $\delta = +k$. Hence the escape-competition effect is particularly strong for such firms. If increases in γ put the industry into states $s = 0$ with greater and greater frequency, then the long-run innovation rate also rises.[18]

The dotted curves in figures 3.11 and 3.12 – corresponding to the fourth panel of figure 3.10 – show that the opposite is true when consumer heterogeneity is high. Then increases in the switching cost push the industry toward $s = 1$ as an absorbing location state. Firms that are *permanently* separated have relatively less convex value functions, since if α is high stage-game profits are non-zero for laggards, and non-negligible for neck-and-neck firms. As a consequence long-run average quality investment is lower than in equilibria where they continually shift locations. By taking the industry toward permanent separation (in the long run), rises in γ therefore reduce the innovation rate.

The distinction between these two cases is striking. It suggests that the long-run

[18]Somewhat confusingly, it is not true that long-run average investment by co-located firms exceeds that by separated firms *in equilibria with continual switching*. If anything it seems that the opposite is true: in the invariant distribution of any such equilibrium long-run average quality investment may be higher in states $s = 1$ than in states $s = 0$. But with fixed locations co-located firms will invest more.

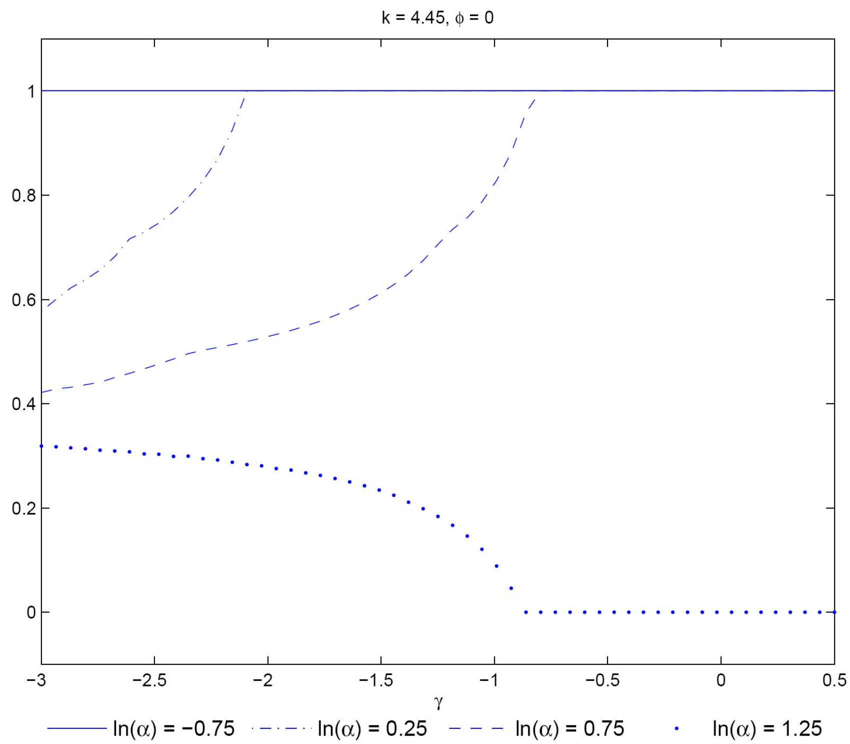


Figure 3.11: Probability that firms are co-located under the equilibrium invariant distribution, as functions of γ

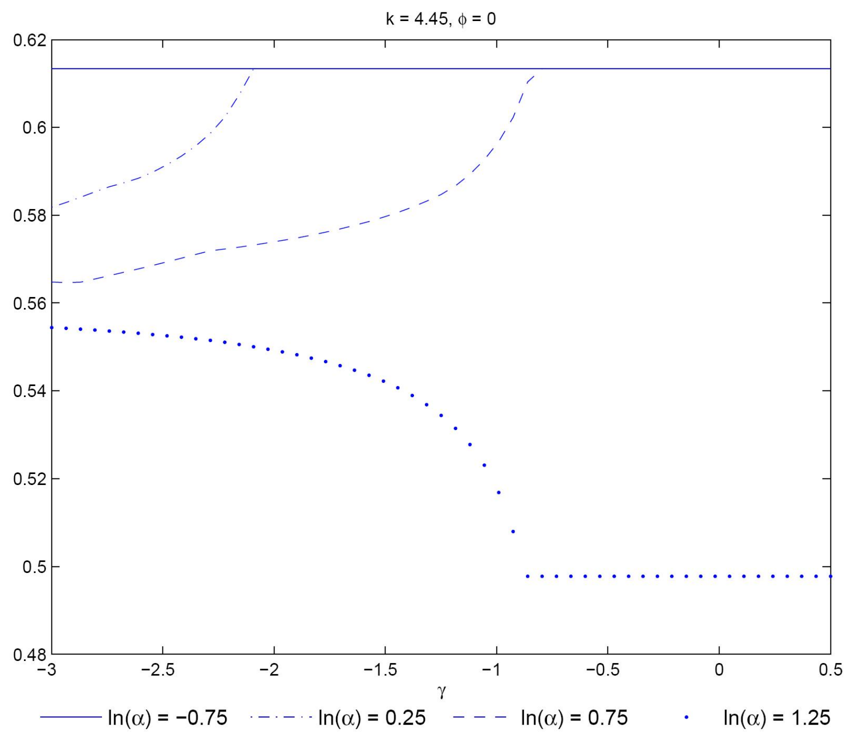


Figure 3.12: Probability of an advance in the frontier technology under the equilibrium invariant distribution, as a function of γ

effects of policies designed to enhance firm mobility depend on, among other factors, the heterogeneity in consumer tastes. Such policies might, for example, include measures designed to lower the costs of entry or adjustment, to facilitate planning of new stores, and so on. The impacts of such 'γ-reducing' measures depends on the degree of heterogeneity in consumer tastes. If heterogeneity is low, then time-averaged horizontal differentiation falls as γ falls, and the innovation rate rises, and *vice versa*. A full welfare analysis of such policies would also need to take into account the impact of changes in firms' horizontal differentiation on consumers, an interesting question which we leave for future work.

Figures 3.13 and 3.14 respectively show the probabilities of co-location, and of an advance in the technological frontier, in the invariant distribution (or distributions), as functions of α. Note that if the switching cost γ is high (=5), then there is no movement at all in equilibrium, in which case there are two invariant distributions, one for the absorbing state $s = 0$ and one for the absorbing state $s = 1$. These distributions correspond to the dotted curves in each figure. For the other values of γ shown there is in equilibrium one ergodic set with a unique invariant distribution.

In figure 3.13 we see the industry transitions alluded to previously in the discussion of figure 3.9. For low α firms see little advantage in horizontal differentiation, and co-location is the unique absorbing location state. At high α firms exploit local market power by operating separately, regardless of their quality differential, and the unique absorbing location state is $s = 1$. As taste heterogeneity rises from the former extreme to the latter, firms enter a region of continual switching – region IV in figure 3.3. They spend less and less time co-located in the long run as the industry passes through this region, eventually exiting into a state of permanent separation. The 'speed' at which the industry passes through region IV is determined by the switching cost γ. When switching is easy a quality laggard starts running away – i.e., investing in horizontal differentiation – even at low levels of taste heterogeneity. Continual switching persists as equilibrium behavior for a relatively broad span of α. On the other hand when switching is costly the industry spends relatively

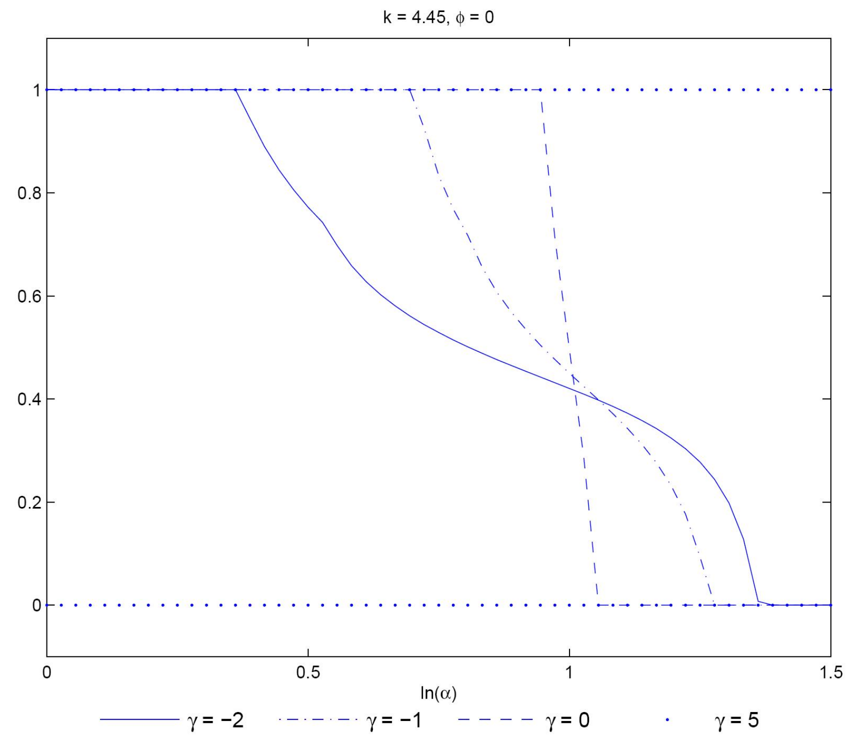


Figure 3.13: Probability that firms are co-located under the equilibrium invariant distribution, as functions of α

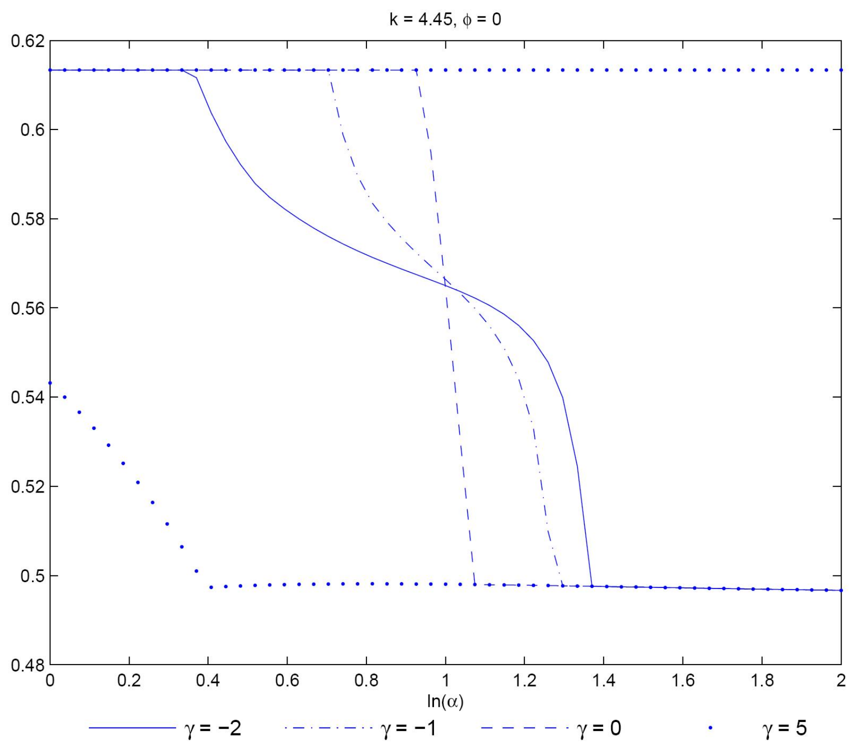


Figure 3.14: Probability of an advance in the frontier technology under the equilibrium invariant distribution(s), as a function of α

little time in continual switching, moving quickly from permanent co-location to permanent separation as α rises.

Effects of α on the industry's long-run innovation rate are shown in figure 3.14. The correspondence between these curves and those for the long-run probability of co-location is again immediately clear. Long-run innovation rates are highest for low levels of taste heterogeneity, because firms are then permanently co-located. As α rises the long-run probabilities of technological advancement and of co-location fall in tandem, eventually levelling off in a state of high taste heterogeneity, permanent separation, and a low innovation rate.

To the extent that α represents firms' degree of market power, figure 3.14 suggests that more competition (lower α) ultimately exerts a positive effect on innovation rates. For small changes in α this need not necessarily be true – recall from figure 3.8 that for some values of ϕ a reduction in α from high levels can cause the rate of innovation to fall initially. However a large enough reduction will eventually set off a transition from more to less differentiation, and firms' long-run average quality investment will accordingly rise.

It is interesting to note that this effect arises in spite of an opposing tendency for firms to spend less time in neck-and-neck competition as α falls. To see this consider figure 3.15, which shows the long-run probability of equal-quality states under the invariant distribution, as a function of α. The effect of more competition (lower α) is negative, because as firms ramp up quality investments they leave the neck-and-neck states sooner. In Aghion et al. (2001) this composition effect allows the industry innovation rate to eventually fall as competition increases. Our model suggests that the countervailing influence of endogenous horizontal differentiation is also important, and in fact may be the dominant effect.

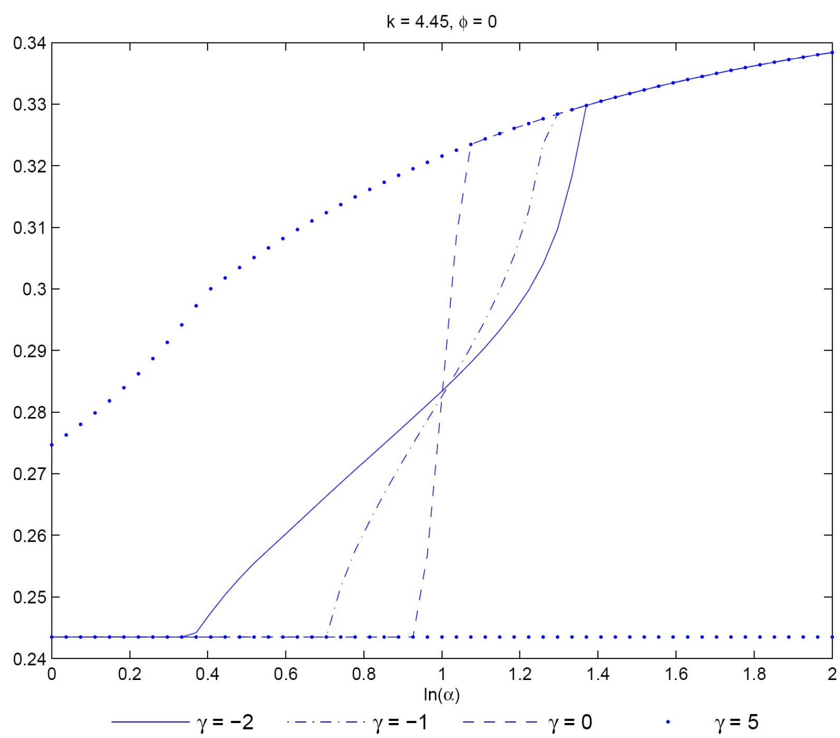


Figure 3.15: Probability that firms are of equal quality under the equilibrium invariant distribution(s), as functions of α

Bibliography

[1] Aghion, Philippe, and Peter Howitt, 1992. "A model of growth through creative destruction,"*Econometrica*, 60(2), 323-51.

[2] Aghion, Philippe; Nick Bloom, Richard Blundell, Rachel Griffith and Peter Howitt, 2005. "Competition and innovation: An inverted-U relationship,"*Quarterly Journal of Economics*, 120 (2) 701-28.

[3] Aghion, Philippe; Christopher Harris, Peter Howitt, and John Vickers, 2001. "Competition, imitation and growth with step-by-step innovation,"*Review of Economic Studies*, 68, 467-92.

[4] Athreya, Kartik, 2005. "Shame as it Ever Was: Stigma and Personal Bankruptcy,"*Federal Reserve Bank of Richmond Economic Quarterly*, 92(2), 1-19.

[5] Ausubel, Lawrence M, 1991. "The Failure of Competition in the Credit Card Market,"*American Economic Review*, 81(1), 50-81.

[6] Aiyagari, S Rao, 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving,"*The Quarterly Journal of Economics*, 109(3), 659-84.

[7] Athey, Susan, and Armin Schmutzler, 2001. "Investment and market dominance,"*RAND Journal of Economics*, 32(1), 1-26.

[8] Barron, John M. and Michael Staten, 2003. “The Value of Comprehensive Credit Report: Lessons from the U.S. Experience,”in *Credit Reporting Systems and the International Economy,* edited by Margaret J. Miller, 273-310. MIT Press.

[9] Bresnahan, Timothy F., and Shane Greenstein, 1999. “Technological competition and the structure of the computer industry,”*Journal of Industrial Economics*, 48(1), 1-40.

[10] Budd, Christopher; Christopher Harris, and John Vickers, 1993. “A model of the evolution of duopoly: Does the asymmetry between firms tend to increase or decrease?,”*Review of Economic Studies*, 60, 543-73.

[11] Burdett, K. and A. Shevchenko. 2006. “Money and the Variety of Goods”, mimeo.

[12] Calem, Paul S.; Michael B. Gorday and Loretta J. Mester, 2006. “Switching Costs and Adverse Selection in the Market for Credit Cards: New Evidence,”*Journal of Banking and Finance* (forthcoming).

[13] Calem, Paul S. and Loretta J. Mester, 1995. “Consumer Behavior and the Stickiness of Credit-Card Interest Rates,”*American Economic Review*, 85(5), 1327-1336.

[14] Camera, G., R. Reed, and C. Waller. 2003. “Jack of All Trades or a Master of One? Specialization, Trade, and Money”, *International Economic Review*, 44, pp. 1275-1294.

[15] Caplin, A. and B. Nalebuff. 1991. “Aggregation and Imperfect Competition: On the Existence of Equilibrium”, *Econometrica*, 59, pp.25-59.

[16] Chatterjee, Satyajit; Dean P. Corbae and Jose-Victor Rios-Rull, 2005. “Credit Scoring and Competitive Pricing of Default Risk, ”mimeo.

[17] Chatterjee, Satyajit; Dean P. Corbae, Makoto Nakajima and Jose-Victor Rios-Rull, 2005. “A Quantitative Theory of Unsecured Consumer Credit with Risk of Default, ”mimeo.

[18] Cooper, R. and D. Corbae. 2002. “Financial Fragility: A Lesson from the Great Depression”, *Journal of Economic Theory*, 107, pp. 159-190.

[19] Corbae, D., T. Temzelides, and R. Wright. 2003. “Directed Matching and Monetary Exchange”, *Econometrica*, 71, p.731-756.

[20] Domowitz, Ian and Robert L. Sartain, 1999. “Determinants of the Consumer Bankruptcy Decision,”*The Journal of Finance*, 51(1), 403-420.

[21] Doraszelski, Ulrich, 2003. “An R&D race with knowledge accumulation,”*RAND Journal of Economics*, 34(1), 20-42.

[22] Doraszelski, Ulrich, and Mark Satterthwaite, 2005. “Foundations of Markov-perfect industry dynamics: Existence, purification, and multiplicity,”mimeo., Harvard University.

[23] Ericson, Richard, and Ariel Pakes, 1995. “Markov-perfect industry dynamics: A framework for empirical work,”*Review of Economic Studies*, 62, 53-82.

[24] Fay, Scott; Erik Hurst and Michelle J. White, 2002. “The Household Bankruptcy Decision,”*American Economic Review*, 92(3), 706-718.

[25] Gross, David B. and Nicholas S. Souleles, 2002. “An Empirical Analysis of Personal Bankruptcy and Delinquency,”*The Review of Financial Studies*, 15(1), 319-347.

[26] Gross, David B. and Nicholas S. Souleles, 2002. “Do Liquidity Constraints and Interest Rates Matter for Consumer Behaviour? Evidence from Credit Card Data,”*The Quarterly Journal of Economics*, 117(1), 149-185.

[27] Grossman, Gene M., and Elhanan Helpman, 1991. “Quality ladders in the theory of growth,”*Review of Economic Studies*, 58, 43-61.

[28] Hotelling, H. 1929. “Stability in Competition”, *Economic Journal*, 39,41-57.

[29] Howitt, P. 2005. “Beyond Search: Fiat Money in Organized Exchange”, *International Economic Review*, 46, pp.405-29.

[30] Huggett, Mark, 1993. “The risk-free rate in heterogeneous-agent incomplete-insurance economies,”*Journal of Economic Dynamics and Control*, 17(5-6), 953-969.

[31] Jaffe, Adam B., and Josh Lerner, 2004. *Innovation and its discontents*, Princeton University Press.

[32] Jovanovic, B. and R. Rosenthal. 1988. “Anonymous Sequential Games”, *Journal of Mathematical Economics*, 17, pp. 77-87.

[33] Judd, K.L., 1998. *Numerical methods in economics*, MIT Press, Cambridge MA.

[34] Kiyotaki, N. and R. Wright. 1991. “A Contribution to the Pure Theory of Money”, *Journal of Economic Theory*, 53, 215-35.

[35] Kiyotaki, N. and R. Wright. 1993. “A Search Theoretic Approach to Monetary Economics”, *American Economic Review*, 83, 63-77.

[36] Kocherlakota, N.R. 1998. “Money is Memory”, *Journal of Economic Theory*, 81, 232-51.

[37] Lagos, R. and R. Wright. 2005. “A Unified Framework for Monetary Theory and Policy Analysis”, *Journal of Political Economy*, 113, 463-484.

[38] Laing, D., V. Li, and P. Wang. 2000. “Money and Prices in a Multiple Matching Decentralized Trading Model”, mimeo.

[39] Langohr, Patricia, 2003. “Competitive convergence and divergence: Position and capability dynamics,”mimeo., Northwestern University KGSM.

[40] Li, Wenli and Pierre-Daniel Sarte, 2003. “The macroeconomics of U.S. consumer bankruptcy choice: Chapter 7 or Chapter 13?”*Federal Reserve Bank of Philadelphia*, Working Paper 03-14.

[41] Livshits, Igor; James MacGee and Michele Tertilt, 2005. "Accounting for the Rise in Consumer Bankruptcies, "mimeo.

[42] Maskin, Eric, and Jean Tirole, 1988. "A theory of dynamic oligopoly, I: Overview and quantity competition with large fixed costs,"*Econometrica*, 56(3), 549-69.

[43] Maskin, Eric, and Jean Tirole, 1988. "A theory of dynamic oligopoly, II: Price competition, kinked demand curves, and Edgeworth cycles,"*Econometrica*, 56(3), 571-99.

[44] Musto, David K. 2004. "What Happens When Information Leaves a Market? Evidence from Postbankruptcy Consumers, "*Journal of Business*, 77(4), 725-748.

[45] Padilla, A. Jorge and Marco Pagano. 2000. "Sharing default information as a borrower discipline device,"*European Economic Review*, 44(10), 1951-1980.

[46] Pakes, Ariel, and Paul McGuire, 1994. "Computing Markov-perfect Nash equilibria: numerical implications of a dynamic differentiated product model,"*RAND Journal of Economics*, 25(4), 555-89.

[47] Reinganum, Jennifer, 1982. "A dynamic game of R and D: Patent protection and competitive behavior,"*Econometrica*, 50(3), 671-88.

[48] Scotchmer, Suzanne, 2004. *Innovation and incentives*, MIT Press, Cambridge.

[49] Shapley, L. and M. Shubik. 1977. "Trade Using One Commodity as a Means of Payment, " *Journal of Political Economy*, 85, 937-68.

[50] Shi, S. 1997. "Money and Specialization", *Economic Theory*, 10, pp.99-113.

[51] Starr, R. and M. Stinchcombe. 1999. "Exchange in a Network of Trading Posts," in G. Chichilnisky, ed., *Markets, Information, and Uncertainty: Essays in Economic Theory in Honor of Kenneth J. Arrow*, Cambridge University Press.

[52] Sullivan, Teresa A.; Elizabeth Warren and Jay Lawrence Westbrook. 2000. *The Fragile Middle Class.* Yale University PRess, New Haven and London.

[53] Whitt, Ward, 1980. "Representation and approximation of noncooperative sequential games,"*SIAM Journal of Control and Optimization*, 18(1), 33-48.

Vita

Borghan Nezami Narajabad was born in Tehran, Iran in 1979. After completing his work at Allameh-Helli High School, Tehran, Iran, in 1997, he entered Sharif University of Technology in Tehran, Iran. He received the degree of Bachelor of Science in Industrial Engineering from Sharif University of Technology in June 2001. In August 2001 he entered the Graduate School of The University of Texas.

Permanent Address: No. 15, 22nd Street
Nasr Place
Tehran, IRAN

This dissertation was typeset with LaTeX 2_ε[19] by the author.

[19]LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay, James A. Bednar, and Ayman El-Khashab.